2809663283

REFERENCE ONLY

UNIVERSITY OF LONDON THESIS

Degree PhD Year 2007 Name of Author SINGLETON, Alexander David

COPYRIGHT
This is a thesis accepted for a Higher Degree of the University of London. It is an unpublished typescript and the copyright is held by the author. All persons consulting this thesis must read and abide by the Copyright Declaration below.

COPYRIGHT DECLARATION
I recognise that the copyright of the above-described thesis rests with the author and that no quotation from it or information derived from it may be published without the prior written consent of the author.

LOANS
Theses may not be lent to individuals, but the Senate House Library may lend a copy to approved libraries within the United Kingdom, for consultation solely on the premises of those libraries. Application should be made to: Inter-Library Loans, Senate House Library, Senate House, Malet Street, London WC1E 7HU.

REPRODUCTION
University of London theses may not be reproduced without explicit written permission from the Senate House Library. Enquiries should be addressed to the Theses Section of the Library. Regulations concerning reproduction vary according to the date of acceptance of the thesis and are listed below as guidelines.

A. Before 1962. Permission granted only upon the prior written consent of the author. (The Senate House Library will provide addresses where possible).

B. 1962-1974. In many cases the author has agreed to permit copying upon completion of a Copyright Declaration.

C. 1975-1988. Most theses may be copied upon completion of a Copyright Declaration.

D. 1989 onwards. Most theses may be copied.

This thesis comes within category D.

- [x] This copy has been deposited in the Library of University College London
- [ ] This copy has been deposited in the Senate House Library, Senate House, Malet Street, London WC1E 7HU.

A SPATIO-TEMPORAL ANALYSIS OF ACCESS TO HIGHER EDUCATION

Alexander David Singleton

A thesis submitted in conformity with the requirements of **Doctor of Philosophy (Ph.D.)**

Department of Geography,
University College London (UCL)

Thesis Submitted August 2007.

Word Count = 75,230.

UMI Number: U593504

All rights reserved

INFORMATION TO ALL USERS
The quality of this reproduction is dependent upon the quality of the copy submitted.

In the unlikely event that the author did not send a complete manuscript
and there are missing pages, these will be noted. Also, if material had to be removed,
a note will indicate the deletion.

UMI
Dissertation Publishing

UMI U593504
Published by ProQuest LLC 2013. Copyright in the Dissertation held by the Author.
Microform Edition © ProQuest LLC.
All rights reserved. This work is protected against
unauthorized copying under Title 17, United States Code.

ProQuest

ProQuest LLC
789 East Eisenhower Parkway
P.O. Box 1346
Ann Arbor, MI 48106-1346

DECLARATION

I, Alexander David Singleton confirm that the work presented in this thesis "A Spatio-Temporal Analysis of Access to Higher Education" is exclusively my own work. Where information has been derived from other sources, I confirm that this has been indicated and cited appropriately.

Alexander David Singleton

ACKNOWLEDGEMENTS

This research was predominantly funded by a Knowledge Transfer Partnership between the Universities and Colleges Admissions Service (UCAS) and University College London (UCL). Two years were spent working in the UCAS Analytical Services Team and specific thanks should be given for the expert programming advice of Nick Edwards and Jim Wilkins. Furthermore, this partnership would not have been created or successfully completed without the input of senior members of UCAS staff including Anthony McClaran, Virginia Isaac and Jill Johnson.

At UCL, gratitude should be extended to David Ashby, Christian Castle, Andrew Crooks, Maurizio Gibin and Pablo Mateos for help, guidance and the provision of a jovial working atmosphere. Specific thanks must be given to Professor Paul Longley who has provided excellent supervision, friendship and support throughout the time completing this thesis.

The final and most important acknowledgement goes to Melissa for supporting, encouraging and enduring me throughout this process, and to my parents who made all of this possible.

ABSTRACT

The purpose of this thesis is to investigate how best to represent the multiple dimensions of social, spatial and temporal processes which shape access to Higher Education in the UK. Full and contextualised understanding of these concepts is seen as acutely important to the stakeholders (pupils, universities, schools) of Higher Education. The agendas of widening participation, extending access and institutional marketing share common challenges to devise better ways of reaching potential students who are appropriately qualified and motivated to pursue and successfully complete the full range of institutional course offerings. One key motivation for this thesis has been the observation that few Higher Education institutions have communication strategies that are tailored towards reaching the full range of potential students who would benefit from their current subject and course offerings. University and college marketing initiatives are often unsystematic, even serendipitous, in the ways in which they identify schools and colleges for outreach and widening participation initiatives, and sometimes uncoordinated in how they present the full institutional profile of subjects of study in these activities. Thus, a core objective of this thesis is to set out some relevant aspects of the changing Higher Education policy-setting arena and to present a systematic framework for widening participation and extending access in an era of variable fees. In particular the thesis aims to illustrate how higher education data and publicly available sources might enable institutions to move from piecemeal analysis of their intakes to institution wide strategic and geographically linked market area analysis for existing and envisaged subject and course offerings.

THESIS OUTPUTS

Refereed Publications

The following two papers are included in the "Appendix – Papers" section:

Longley, P., Singleton, A.D., (2008) Classification through Consultation: Public Views of the Geography of the e-Society. *International Journal of Geographical Information Science*. Forthcoming.

Longley, P.A., Singleton, A.D. (2008) Regional Variation in Material Deprivation and e-Engagement. *Papers in Regional Science*. Forthcoming.

Refereed Conference Proceedings

Singleton, A.D., Farr, M. (2004) Widening Access and Participation in Higher Education. *Proceedings of the GIS Research UK 12th Annual Conference*. 28th - 30th April 2004. Norwich, University of East Anglia.

Singleton, A.D. (2005) Widening Access and Participation in Higher Education. *Proceedings of the GIS Research UK 13th Annual Conference*. 6th - 8th April 2005. Glasgow, University of Glasgow.

Singleton, A.D. (2006) Bespoke Versus General Purpose Discrete Classifications: Segmentation of Higher Education Market Data. *Proceedings of the GIS Research UK 14th Annual Conference*. 5th-7th April 2006. Nottingham, University of Nottingham.

Singleton, A.D., Longley, P. (2007) Modifying the 2001 Census Output Area Classification for Applications in Higher Education. *Proceedings of the GIS Research UK 15th Annual Conference*. 11th – 13th April 2007. Ireland, NUI Maynooth.

Contributions to Books

"The Geodemography of Geography" pp 483 in Geographic Information Systems and Science (Second Edition). Chichester, Wiley (P A Longley, M F Goodchild, D J Maguire, D W Rhind).

Working Papers

Singleton, A.D. (2004) A State of the Art Review of Geodemographics and their Applicability to the Higher Education Market (online). CASA Working Paper 74. London: CASA. Available from http://www.casa.ucl.ac.uk/publications/workingPaperDetail.asp?ID=74.

Singleton, A.D., Davidson-Burnett, G., Longley, P.A. (2007) University Market Area Analysis for Widening Participation (online). CEBE Working Paper. 12. Available from: http://www.cebe.heacademy.ac.uk/publications/stats.php?file_name=publications/workpapers/pdf/WorkingPaper_12.pdf&resources=WorkPapers

Other Conferences and Presentations

Singleton, A.D. (2003) Tailored Geodemographics - The Construction of an Educational Mosaic. *CASA Seminar Series*. 12th November 2003. London, CASA.

Singleton, A.D. (2003) Geodemographics and Higher Education. *ESRC Research Methods - New Representations: The Use of Geodemographic Classifications in Research and Public Service Delivery*. 18th -19th February 2004. London, CASA.

Singleton, A.D. (2004) Applying Geodemographics to Education. *New Representations: The use of Geodemographic Classifications in Research and Public Service Delivery*. 18-19th February 2004. London, CASA.

Singleton, A.D. (2004) A Meritocratic Market? The Geographies of Access and Retention in Higher Education. *CASA Seminar Series*. 9th December, 2005. London, CASA.

Singleton, A.D. (2005) The Social Space of UK Higher Education Acceptances. *Royal Geographical Society Conference*. 31 August – 2 Spetember. London, Royal Geographical Society.

Singleton, A.D. (2006) Under the Spotlight. Geodemographic Segmentation of Higher Education Market Data. *Regional Science Annual Conference*. 16th – 18th August 2006. Jersey, Regional Science Association.

Longley, P., Singleton, A.D. (2006) Circling the Drain. *Royal Geographical Society Conference*. 30th August – 1st September 2006. London, Royal Geographical Society.

Singleton, A.D. (2006) Higher Education Segmentation. *Royal Geographical Society Conference*. 30th August – 1st September 2006. London, Royal Geographical Society.

Singleton, A.D. (2006) – Demonstration of Institutional and Subject Geodemographic Profiling. *Spatial Literacy, Student Recruitment and Widening Participation* 12th June 2006. London, University College London.

Invited Conferences and Presentations

Singleton, A.D. (2004) Geodemographics for Education. *UCAS Admission Officers Conference*. 5th – 7th April 2004. Reading, University of Reading.

Singleton, A.D. (2004) Tailored Geodemographics - The Construction of an Educational Mosaic. *LSDA Research Conference*. 8th - 9th December, 2005. London, Paddington Hilton.

Singleton, A.D. (2005) Marketing a New Market? *Admission Officers Conference*. 12th – 14th April 2004. Keele, Keele University.

Singleton, A.D. (2007) *A Meritocratic Market? Tailoring Geodemographics for use in Higher Education Widening Participation*. January 24th, 2007. Bristol, Bristol University.

Singleton, A.S. (2007) Open Source Geodemographics: Applications for Education. 22nd March 2007. *Sheffield, ESRC Research Methods – Geodemographics and the Social Sciences*. Sheffield, The University of Sheffield.

Longley, P.A., Singleton, A.S. (2007) *Turning Data into Information: Geodemographic Applications*. 14-18th May 2007. Koyoto, Ritsumeikan University.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

PART 1: WIDENING ACCESS? DATA, TOOLS AND TECHNIQUES

1

INTRODUCTION

1.1 Personal Motivation

The motivation for this thesis is embedded in previous personal experiences and sustained by a desire to contribute a body of work which appropriate organising bodies could use to create real advantages to wider society. Before introducing the structure and topics investigated by this thesis those personal motivations which have led to this research are outlined. The author of this thesis comes from a grammar school educated family where participation in Higher Education was a natural progression in life course from school to adulthood, where attendance at a "good university" was assumed and understood by all as valuable for both personal and career development. The benefits placed on education were always made explicit, be they the provision of skills which equip one for a successful career, or the intangible experiences and lifelong friendships which resulted from attendance. Good schooling was provided through living in reasonably affluent residential areas in close proximity to the best local state schools, and when applications to Higher Education were made these were guided towards those traditional institutions familiar to the author's parents. In general terms good schooling and guidance are not, however, meritocratic, and the author of this thesis feels in a privileged position to now be able to contribute to a body of research which aims to improve equality of the opportunities to which he has been privy.

The themes developed in this thesis were introduced to the author while studying as an undergraduate at the University of Manchester where a dissertation project was conducted to look at issues of widening participation in the university. This exploratory study highlighted stark contrasts in Higher Education participation rates between the most and least socio-economically advantaged members of society and highlighted how the then University of Manchester had a strong regional recruitment market. At the time geographical research in this area was relatively new, but since this time widening participation in Higher Education has grown into an increasingly significant area of public policy, viewed as imperative to improving social mobility amongst hitherto disadvantaged groups and to educating the knowledge workers that are pivotal to a sustained and internationally competitive economy.

The author was given the opportunity to work with UCL and the University and Colleges Admissions Service (UCAS) on a two year joint project where themes of spatial and social exclusion were investigated using UCAS applicant data. These analyses were conducted as part of the analytical services team which process user requests for data and analysis. During this period the author became aware that many analyses requested within the sector, that were used to make appropriate decisions relating to extending participation in Higher Education were unsophisticated and often flawed. Furthermore, there was a lack of centralised provision of simple tools to convert very rich data into the information necessary to make appropriate decisions, and that as a result a range of private sector organisations were filling these analytical roles at greater cost than was necessary.

It is within this personal history which the author positions his thesis as an academic contribution to the literature on widening access and as a practical contribution which suggests how knowledge gained can be transferred and implemented in national data services to better meet the needs of stakeholders in Higher Education.

1.2 Problem and Approach

The core problem investigated by this thesis is a consideration of how access to Higher Education can be made more equal between society's constituent groups. The approach taken by this thesis to meet these ends broadly follows the framework presented by Longley *et al* (2005) in Figure 1.1.

Figure 1.1: A Decision Making Infrastructure

Empirical facts about raw geographical and social processes on education participation can be assembled and structured to enable evidence to be created. When this evidence is linked to theory presented in the literature plausible explanation can be made as to why particular processes or behaviours are occurring. Using the knowledge gained from these analyses, the findings can be extrapolated to more general statements about the functioning of society, and through developed wisdom be implemented in practical solutions which should enable equality of access to Higher Education to be targeted for improvement.

Key to this exercise is the creation and use of an organising framework, which in this thesis is created by using geodemographic data. Geodemographics has been defined by the analyst Peter Sleight (1997, p 16) as the "analysis of people by where they live". These data can reveal social, economic and demographic conditions that characterise areas as small as the unit postcode. A range of these classifications are used and evaluated in this thesis in the interests of impartiality and comprehensiveness.

1.3 Structure and Content

This thesis consists of 10 chapters which are divided into four parts:

- Widening Access? Data, Tools and Techniques

- Geodemographic Profiling and Decision Support for Higher Education
- Geodemographics for Higher Education
- Are Things Getting Better?

The first part of this thesis will describe the development of the contemporary Higher Education system through the historical changes in participation and access rates. The current data collection and organisation mechanisms within the sector will be described to highlight overlaps, problems of coordination and missing data. It will also be shown how despite a growing demand and requirement for information which integrate these disparate datasets, these are not currently being provided through centralised services to key stakeholders in Higher Education. Methods of social measurement will be introduced, and it will also be suggested how concepts of social capital can map into the quantitative variables available which permit generalisations.

The second part of this thesis will demonstrated how stakeholders in Higher Education are currently making uninformed decisions in areas of policy concern through a lack of appropriate decision support tools. This will begin with an overview of Higher Education spatial and social complexities, including detailed description of a number of variables including distance, type of institution, tariff score and course choice. A pilot decision support tool will be presented which will link a user requirements analysis with theoretical concepts and empirical observations to demonstrate a unique methodology for the integration of data from the Higher Education and school sectors.

The third part of the thesis will extend the concept of classification to examine how more relevant indicators for Higher Education applications can be built and evaluated. This will challenge the implied assumption held by commercial classification builders that an individual's use of public services is analogous to the ways in which consumers use private goods.

The final part of the thesis will examine whether Higher Education inequalities are getting better or worse and how these may be addressed through engaging with stakeholders in Higher Education.

2

A MERITOCRATIC MARKETPLACE? THE GEOGRAPHIES OF ACCESS AND RETENTION IN HIGHER EDUCATION

2.1 Historical Development of Access and Participation Inequalities

Inequality of participation and access to Higher Education has a long history extending back to medieval times. The interplay between external influences such as industrialisation, and centralised policy change has shaped spatial and societal Higher Education access patterns. To understand the heterogeneous nature of contemporary Higher Education, it is important that these historical patterns of access be identified.

Towards the end of the 12th century the English Universities of Oxford and Cambridge were established. Cobban (1999) discusses that before the 16th century those students attending institutions were of middle to lower social condition, and were structured less on wealth or class, and markedly removed from the rigidity of the social structures outside academia. However, the system became more elitist towards the end of the 15th century with an influx of students from noble births. This contrasted with the acceptance patterns of European institutions that throughout the medieval period had a considerable number of noble scholars. Participation was limited during the medieval period with only 3% of the male population attending university (Ainley, 1994).

The Civic Universities were established during the 19th century as a solution to the growing educational demands of rapidly developing industrial cities. These institutions developed a very different pedagogy to the medieval institutions, with curriculum centred on meeting the needs of a newly growing industrial society. Table 2.1 plots the establishment of the Civic Universities

Table 2.1: The Development of the First Phase Civic University (Adapted from Ross, 2003).

University	Date	Notes
Durham University	1833	
London University	1836	Amalgamation of Kings College and University College which were established 10 years earlier.
Victoria University	1880	Constituted of Colleges, Manchester, Leeds and Liverpool. These split in early 1900s.
Birmingham	1900	
Sheffield	1905	
Bristol	1909	

Most of these first phase Edwardian 'redbrick' institutions originated from older institutions such as working men's colleges or institutions (Ross, 2003). However, the second wave of institutions established close to WWI had only some of their origins in these institutions. Originally these institutions would teach University of London Degrees, but gradually were granted independent status (Ross, 2003). These institutions included Reading, Hull, Nottingham, Southampton, Exeter and Leicester.

Anderson (1992) describes the class structure in the Civic University System as following three distinctive phases of development. During the early phase the institutions continued the traditional task of serving the older land owning professional elite. However, during the mid phase around 1860 these entry profiles began to change with institutions adapting to the needs of the industrial society with an increasing class mix. The late phase proceeded with a shift towards a more diverse and middle class professional system, leading to a faster increase in student numbers, and a greater expansion as middle class occupations grew. The changes occurring in the civic system challenged the traditional universities of Oxford and Cambridge, therefore forcing them to reform between 1850 and 1870.

After World War II there were various Higher Education reforms with the aim of treating all scholars equally, achieved through mandatory grants, and a selection process for university based upon the Advanced Level and Scottish Higher examinations.

UCCA (1994) discusses how two major concerns dominated university admission staff during the 1950s. Firstly, applications for undergraduate courses were tending to increase year by year as more young people were entering sixth forms. A second concern was demographic, with the expectation that the university population was about to increase rapidly as the post war baby boom reached Higher Education participation age. UCCA (1994) quantifies this growth in applications between 1955 and 1960 with Nottingham University reporting an increase of 150% and Leeds University 130%. This trend was boosted by individuals applying to more than a single university, which occurred to a lesser extent before WWII. Ainley (1994) observes that Higher Education had grown to 7.2% of the age cohort by 1962.

In 1961 the Universities Central Council on Admission (UCCA) was set up to as the centralised admissions agency for UK courses of full time Higher Education. The key aims were to create a fairer and more objective system of entrance that would also reduce the administrative burden on individual institutions. It should be noted that when Polytechnic Colleges were established during the 1980s, the Polytechnic Central Admissions Service (PCAS) rather than UCCA managed their applications. Figure 2.1 shows how during the period 1962 to 1990 there were a steady increase in both home and overseas applications and acceptances.

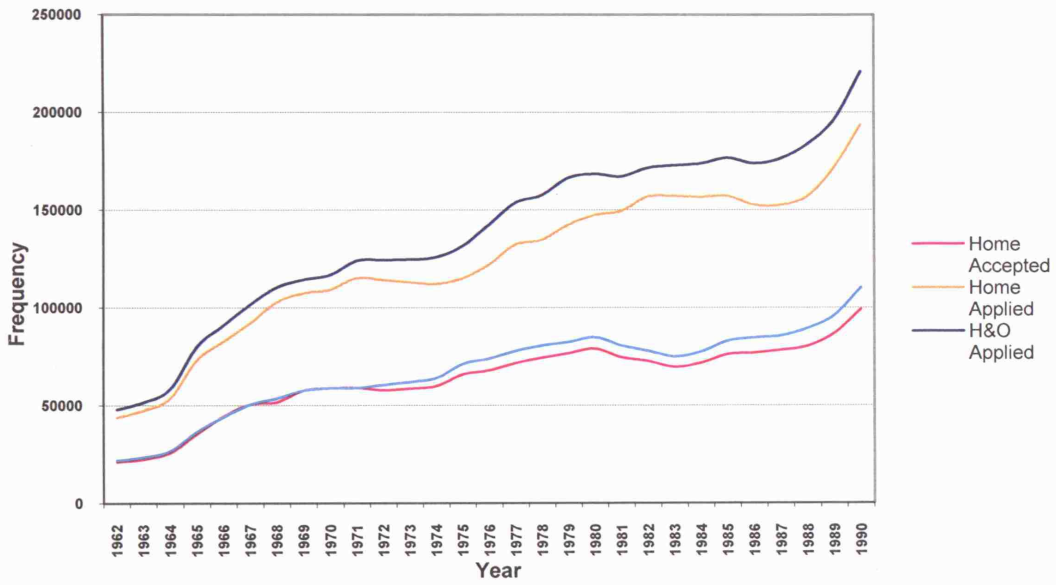


Figure 2.1: Applications and Acceptances 1962 – 1990 (Source: UCAS)

In 1963 a report compiled under Lord Robbins (Robbins, 1963) set out recommendations that courses of Higher Education should be made available to all of those who were qualified by attainment to pursue them and who wished to do so. It established that the majority of this Higher Education would be based in universities, although it also observed that there was also a growing need for more vocational education. In 1966 a White Paper entitled 'A Plan for Polytechnics and Other Colleges' was delivered in parliament (Secretary of State for Education and Science, 1966). In this report it was suggested that:

> *The Government have committed themselves to an even greater expansion of higher education than was forecast in the Robbins report, and in this document announce their intention of developing a new sector of higher education within the Further Education system, to complement the universities and colleges of education… The generic term for these new centres is to be 'polytechnics'.*

Secretary of State for Education and Science (1966:2)

These recommendations established a binary divide between traditional universities and the new type of institution called polytechnic colleges (Ainley, 1994). Polytechnic colleges were established to push comprehensive principles into Higher Education by offering a broader range of opportunity to a wider base of students, feeding local community needs through local education authority control, and eroding 'the elitist character of British HE' (Carr, 1998: 275). The rationale for polytechnic education can be traced back to the writings of Marx and Engels, where real-life problems are tackled and reflected on jointly by teachers and the taught. Ainley (1994) discusses that the 'universities for all' principle pioneered what the Polytechnic's idealistic founders referred to as liberal vocationalism, and that this shift in pedagogy reflected changes that had occurred in contemporary comprehensive and community schools. The polytechnic education aimed to make available opportunities to qualify for occupations on equal terms with those educated selectively. Thus, the polytechnic system began to widen access with the percentage of full and part time students growing to 12% of the age range by the end of the 1970s (Ainley, 1994).

During this period there was a significant shift in policy towards reshaping Higher Education under conditions of severe resource restraint, beginning with publication of the 1981 Public Expenditure White Paper (Eurydice, 2004). This stated:

> ***"This is likely to oblige institutions to review the range and nature of their contribution to higher education. It is also likely to lead to some reduction in the number of students admitted to higher education with increased competition for places"***

This resulted in a reduction in public sector expenditure in Higher Eucation of around 8% over 3 years. The rationalisation of Higher Education continued until around 1987 when the publication of the White Paper 'Higher Education: Meeting the Challenge' set out policy changes to increase participation rates and widen access to Higher Education for non traditional applicants, such as mature students or those without A-Levels (Eurydice, 2004).

The binary divide between the two systems had become blurred with Polytechnics moving towards the academic norms of universities, while universities were becoming less elitist and inward looking. Jenkins (1995) argued that 'polytechnics had not become universities, universities had become polytechnics'.

Another White Paper was released in 1991 entitled 'Higher Education, A New Framework' (Department for Education, 1991), which set out proposals to remove the binary divide between universities and polytechnic colleges, and was passed through parliament as part of the 1992 Higher Education Act. Carr (1998) describes the three key aims of this post-binary policy as:

- Creating funding agencies that would be more sensitive to the secretaries of state rather than acting as buffer between academics and the state.
- Continuing moves towards a competitive funding system, where former polytechnic colleges could compete for a share of £680m allocated to universities for research.
- Create a dual system of quality assurance incorporating the old universities, and in time generate league tables to underpin student choice in the new unified Higher Education system.

This influx of institutions and students into the Higher Education system, in a decade of underfunding has led to large deficits in the finances of many of the UK's leading institutions. Funding has been improving in recent years, however in 2002/03 there were still 48 institutions in deficit with the largest £6.2m (MacLeod, 2004). Institutions created numerous coping strategies to deal with these debts such as increasing the number of student places. The fall in public funding per student is shown in Figure 2.2, first published in Dearing (1997).

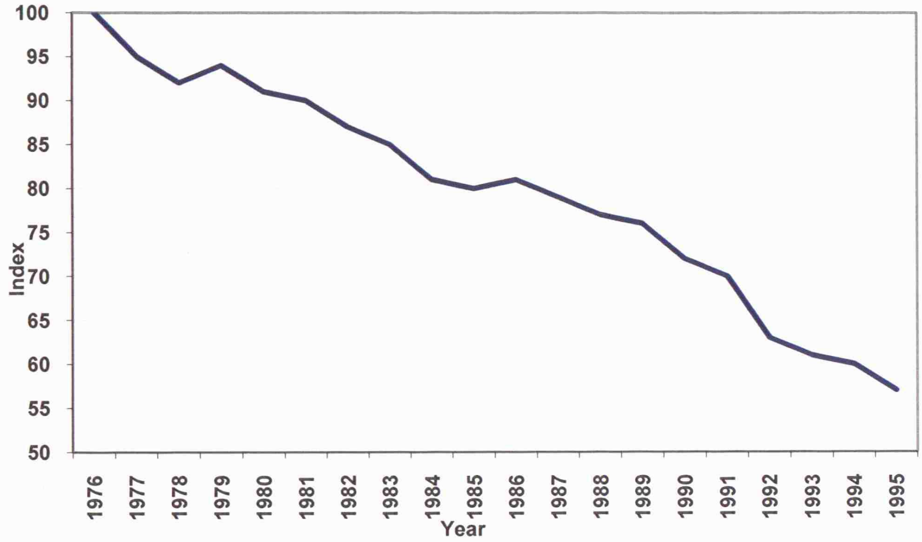


Figure 2.2: Index of Public Funding per Student in Higher Education 1976-95 (Dearing, 1997)

In line with the post-binary policy changes PCAS and UCCA were amalgamated during 1993. 1994 entry was managed by a new organisation, the Universities and Colleges Admissions Service (UCAS). The effect of this amalgamation can be seen in the apparent massive upturn in both applications and acceptances during the 1994 recording period (see Figure 2.3). There is a second increase around 1997 where the Art and Design Admissions Registry (ADAR) was amalgamated with UCAS. Post 1998 there is a continual and steady growth in the frequency of both applications and acceptances.

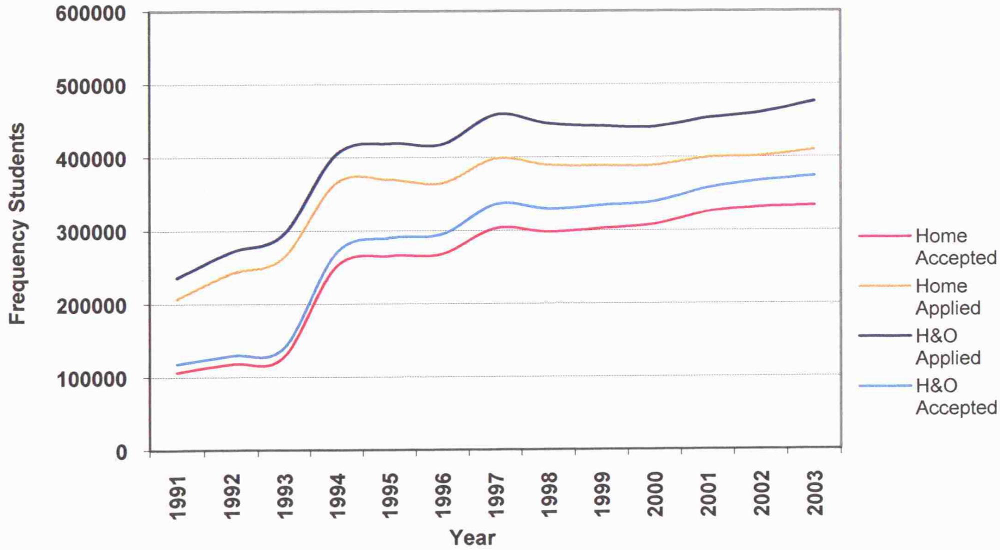


Figure 2.3: Applications and Acceptances 1991 – 2003 (Source: UCAS)

When considering participation rates it is important to understand how the addition of institutions to what is being classified as Higher Education can affect these rates. Figure 2.4 below shows how the number of UCCA and UCAS institutions has changed dramatically since 1963.

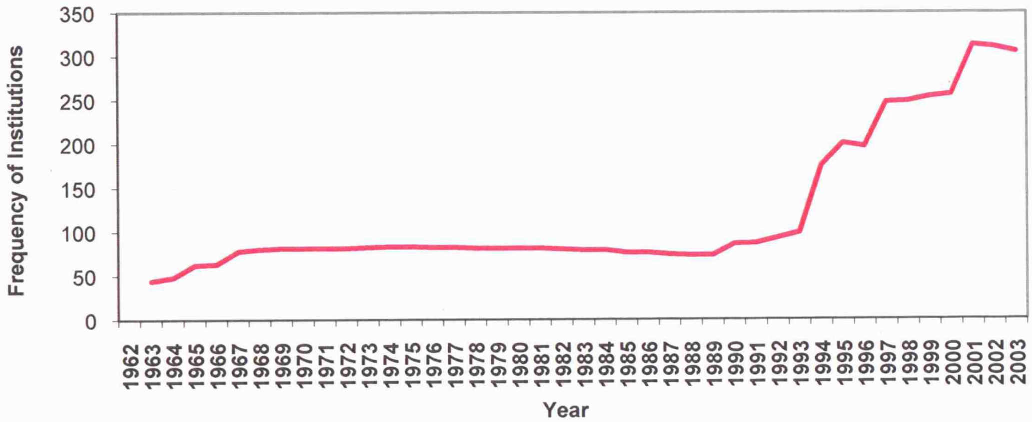


Figure 2.4: UCCA and UCAS Institutions 1962 – 2003 (Source: UCAS)

The Dearing Report (1997) investigated how the purpose, shape, structure and funding of Higher Education should meet the needs of the United Kingdom for the following 20 years. This report was far reaching in its objectives and included recommendations about:

- Demand for Higher Education
- Widening Participation
- Students and Learning
- Content of Programmes
- Qualifications and Standards
- Research and Funding
- Regional and Local Impact of Institutions
- Human Resources Issues
- Funding

In 1997 the newly elected Labour Government acted on the Dearing Report with recommendations to scrap the remaining grants and replace them with student loans linked to parental income. There was also the announcement that means tested tuition fees would be introduced to the sum of £1000.

The Dearing Report highlighted issues of *access* rather than *participation*, discussing that there remain groups in the population who are underrepresented in Higher Education including lower socio-economic groups and certain ethnic minorities. Participation and access are subtly different terms, both of which will be used throughout this thesis. In public policy discussion they are often used interchangeably, however in this thesis participation will refer to increasing the absolute frequency of students attending Higher Education, whereas access refers to readdressing the balance between the constituent groups making up the total participation.

Since the Labour Party was re-elected in 2001, an ambitious 50% Higher Education participation target has been set. Although this target has been debated as arbitrary (May, 2003), recently released figures suggest current participation rates have reached 45% of the relevant age cohort (Economist, 2004). However, the extent to which this may be apportioned between demographic change, improving A-Level performance and widening participation initiatives is debatable.

Many institutions now face financial crisis because funding has not kept pace with increasing student numbers. Funding in Higher Education has fallen in real terms by 38% since 1989, resulting in massive underinvestment (Brown and Piatt, 2001). The Oxford Centre for Higher Education Policy Studies (OxCHEPS) estimates the costs to educate a 2003 student at Oxford University to be around £18,600 per annum, where only 6% of this cost comes from student fees. Of the remaining 94% around half comes from private sources such as endowments, with the government contributing the other half through the Higher Education Funding Council for

England (HEFCE) (Palfreyman, 2004). Perhaps of most concern was the following key findings arising out of comparison between Oxford University and leading top-tier public and private US universities:

- There is considerable relative under investment in infrastructure and other academic support structures in Oxford.
- The expenditure per undergraduate is three times higher in the US elite universities of Harvard and Princeton.
- Academic salaries are one-third, to one-half of those at top U.S. universities

The re-election of the Labour Party in 2001 came with a manifesto promise stating 'we will not introduce 'top-up' fees and have legislated to prevent them' (Blair, 2000). In 2003 the White Paper 'The Future of Higher Education' was published. This paper had two interrelated purposes: to outline the funding gap required to maintain international teaching and research standards; and also to create conditions for equality of access. A bill containing the recommendations from this paper was narrowly passed (by 5 votes) through parliament in 2004. The changes included:

- A capped variable tuition fee up to the value of £3000.
- Establishing an access regulator.
- Introducing maintenance grant for poorest 30% from £1000 - £1500.
- Poorest students receive fee remission for the first £1500 (£1200 by state and £300 by institution).
- Student loans to be aligned with the real cost of living.
- All student debt dropped after 25 years.

The previous discussion has showed that increasing participation has almost certainly occurred, however the extent to which it has "widened" is debateable (Farr, 2002). UCAS statistics show that from 1996 to 2002, home applicants for full time degrees rose by 26% (UCAS, 2003a). However, if the Mosaic lifestyle group indices of these applications are viewed it can be seen that representation is not even across all geodemographic types. In the year 2000 the High Income Families group is over represented with an index of 200 (100 is average), whereas Council Flats are indexed at 49 and low rise council flats at 48 (UCAS, 2001). It is these inequalities that have raised questions surrounding social justice and discrimination within the system.

If the acceptances to both UCCA and UCAS are examined by the National Statistics Socio Economic Classification (NS-SEC) and previous equivalents such as Occupational Group and Social Class it is possible to see that the access patterns between groups have been relatively static throughout the last 40 years. However, it should be noted that temporal analysis of UCCA / UCAS data is with the caveat that the institutions from which the total population is derived do not remain static, with some institutions entering, leaving and amalgamating each year. The effect of these changes on the total intake of institutions was show earlier in Figure 2.4.

During the period 1968 – 1978 Occupational Group (Rose, 1995) was used to classify accepted students parental occupation. The pattern is relatively static, perhaps with a slight increase of Professional & Technical occupations and a decrease in Manual (See Figure 2.5).

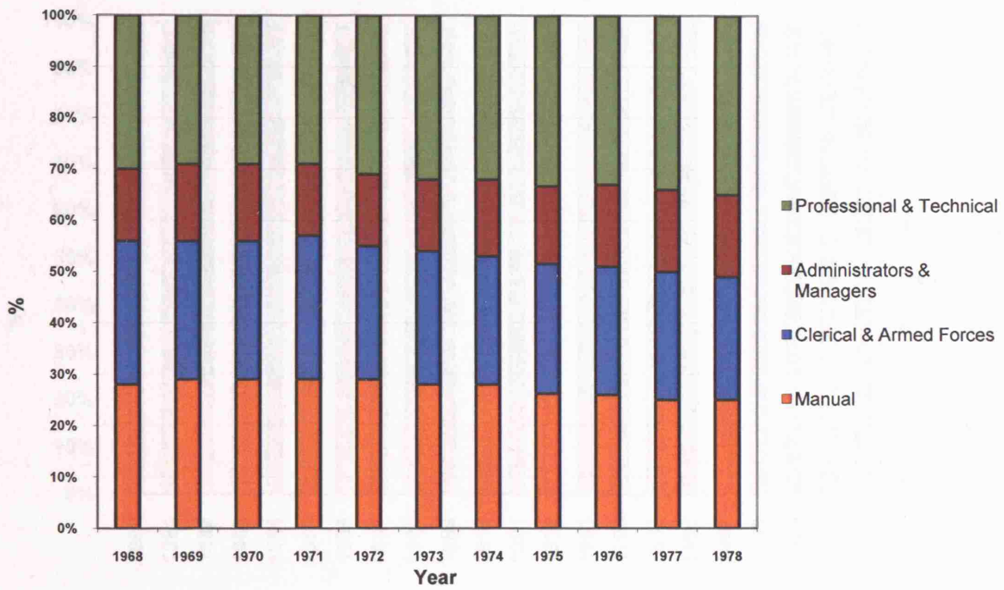


Figure 2.5: Occupational Group 1968 - 1978 (Source: UCAS)

Figure 2.6 shows the change in the Social Class (Rose, 1995) of accepted students between 1980 and 2001. The 1980-1993 data were extracted from UCCA statistical bulletins, and these data do not record students with parental occupation classified as unknown, however these students are recorded for UCAS acceptance during the period 1994 – 2001, and this accounts for the sudden change in profile between 1993 and 1994. In the 1980 – 1993 period the access rates remained relatively static with professional and intermediate social classes dominating acceptances. During the period 1993 – 2001, once the Polytechnic and University admissions systems had been combined, a slightly larger percentage of total acceptances derived from the lower social class groups. Also during this period there was a growing number of unknown classifications, perhaps as the Social Class schema was becoming progressively out dated.

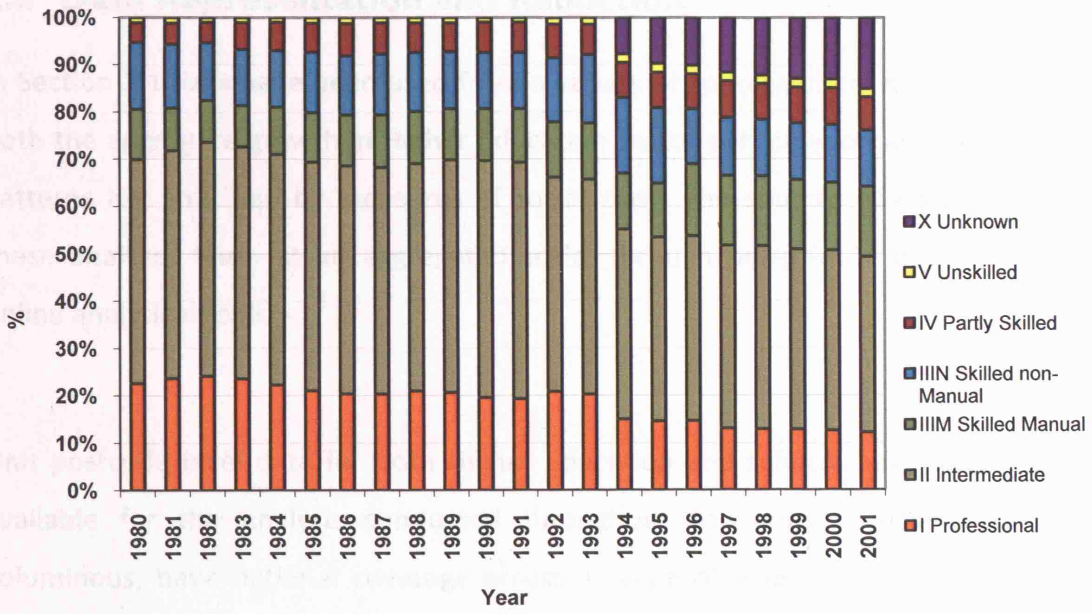


Figure 2.6: Social Class 1968 – 1978 (Source: UCAS)

Finally, in 2002 a new Socio-Economic Group classification[1] was adopted. From looking at these proportions, it appears that despite government intervention the socio-economic profile of acceptances remains reasonably static, perhaps with some growth in the frequency of students classified as "Unknown" (See Figure 2.7). This theme is investigated further later in the thesis.

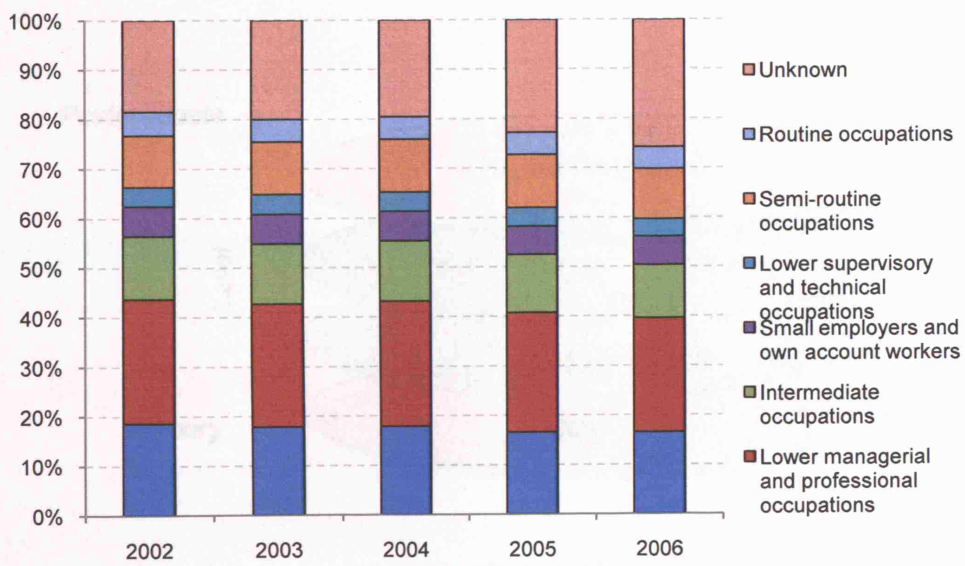


Figure 2.7: Social-Economic Group 2002-2006.

[1] http://www.statistics.gov.uk/methods_quality/ns_sec/cat_subcat_class.asp

2.2 Data Representation and Reduction

In Section 2.1 data have been used from a variety of sources to create a picture of both the aggregate growth in Higher Education sector participation and how these patterns are stratified by measures of social class. The sources of data used for these analyses were at an aggregated scale, through pre-printed publication or online analytical tools.

Unit postcode level data for both Higher Education and schools have been made available for the analysis conducted throughout this thesis. These data are voluminous, have national coverage across a range of levels, but do not have a common identifier linking individuals through time. Section 2.2.1 describes how the various educational datasets are created, and which organising bodies are custodian to this information.

2.2.1 Data Management in UK Education

The UK education system offers a number of routes and potential outcomes in terms of qualification type, progression routes and organising bodies (See Figure 2.8). The system is divided into a series of levels from Primary through to Postgraduate Degree and at present is compulsory until the age of 16.

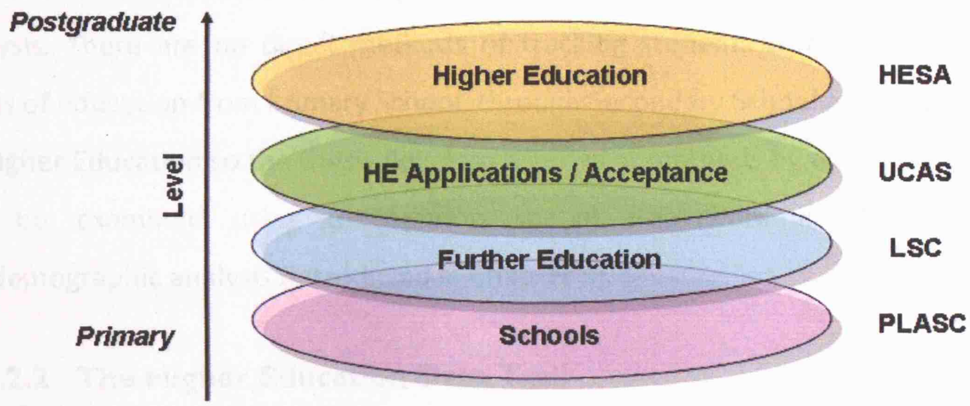


Figure 2.8: An Overview of the State Education System

Assembly and interpretation of educational statistics is difficult because of the plethora of definitions of educational qualifications, the overlapping responsibilities of data collecting organisations, and the inability to track the educational careers of individuals through unique identifiers. The analysis in this thesis uses various national Higher Education and FE datasets that are collected by a number of different agencies. These are summarised in Table 2.2.

Table 2.2: Educational Data Collection Agencies

Sector	Data Sets	Agency	Contents
FE	Individualised Learner Record (ILR)	Learning and Skills Council (LSC)	Data relating to English post-16 education and training.
FE/HE	Courses, Applicants and Acceptances Database	Universities and Colleges Admissions Service (UCAS)	UK and international dataset holding information on UCAS applications, offers and their outcomes.
School	Pupil Level Annual School Census (PLASC)	Department for Education and Skills (DfES)	UK dataset relating to individual students in all schools.
HE	Student; Student First Destinations; Staff; Finance and the Non-credit-bearing Course Records	Higher Education Statistics Agency (HESA)	Various UK datasets covering various aspects of staff and students in Higher Education.

These data contain records of individuals that each include a unit postcode level attribute and offer national coverage across a range of years. There is no single UK source of education data and access was negotiated with each separate provider for purposes of this research. Description of those data used will accompany each analysis. There are no direct methods of tracking students through the various levels of education from Primary School, through Secondary School and into Further or Higher Education so the thesis develops a series of methods by which the sectors can be examined using a common spatial framework in the form of geodemographic analysis (introduced in Chapter 3).

2.2.2 The Higher Education Data Trail

Higher Education data collection and dissemination are not managed by a single organisation. These data are predominantly created through the applications process to full time Higher Education and involve a range of transactions between individuals and national organisations. A simplification of how the admissions process creates a range of data is shown in Figure 2.9.

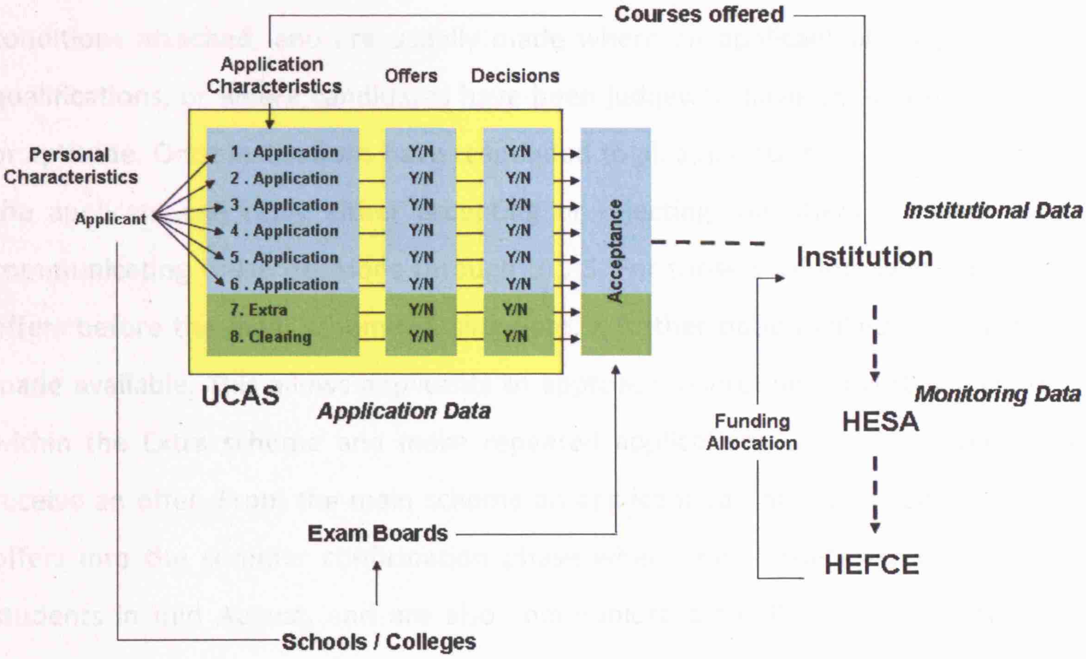


Figure 2.9: A Simplified Model of Higher Education Data Flows

The Universities and Colleges Admissions Service (UCAS) centrally manage the application process for all full time courses of Higher Education in the UK. The UCAS nomenclature defines 'Applicants' as those individuals seeking entry to Higher Education through UCAS. Applicants to Higher Education make an initial selection of six choices (applications) which each consist of an institution, course and campus selection. UCAS collect these data alongside various other attributes of the individual applicant, e.g. age, gender, address etc. The majority of applicants now submit their applications electronically and much of the data processing is automated. For the raw UCAS database to remain accurate, this requires applicants to have both submitted data truthfully and also with minimal error. Where errors are detected though automated checking, these are flagged for human intervention, and data processing operatives correct these details where possible. The main application cycle begins in October of each year and runs through to a deadline at the end of June in the following year. During this period applicants receive offers or rejections from their six or fewer applications. These decisions are communicated on behalf of the institutions through UCAS, as conditional offers, unconditional offers or rejections. Conditional offers usually specify a particular grade attainment or subject requirement. Unconditional offers are those with no conditions attached, and are usually made where an applicant already has prior qualifications, or where candidates have been judged to have exceptional promise or aptitude. Once institutions have responded to all applications from an applicant, the applicant can reply either accepting or rejecting the offers received, again communicating these decisions through UCAS. For those students who receive no offers before the main scheme closing date, a further option called UCAS Extra is made available. This allows applicants to approach institutions advertising courses within the Extra scheme and make repeated applications until they successfully receive an offer. From the main scheme an applicant can hold a maximum of two offers into the summer confirmation phase when exam results are delivered to students in mid August, and are also communicated to UCAS. This allows those applicants with open conditional offers to be automatically confirmed or declined based on a comparison between attainment and offer requirements. Those applicants who apply after the main application scheme closing date, those who do not receive offers from any institutions, and those who do not meet their offer requirements are placed into a process called clearing, where institutions advertise their remaining course places. Negotiation then occurs between applicants and institutions directly, and if places are found, these are again communicated through UCAS. It is through this yearly application cycle that the majority of Higher Education data are collected for the UK. Institutions are supplied electronically with data collected by UCAS for their admitted applicants, and these in turn populate internal acceptance databases or student records. Once these data are within the institutions they can be updated or amended as necessary, although such subsequent changes are not returned to UCAS. For undergraduate admissions, UCAS only collect data on full time admissions, and only from those institutions within the UCAS scheme. This currently excludes the University of Buckingham, Birkbeck College (University of London) and the Open University. The UCAS scheme also includes a wide range of smaller colleges offering Higher Education level full time qualifications, often referred to as university sector colleges.

Institutions are funded through a mixture of public funding through central government, private enterprise or investment, entrepreneurial activity and interest from endowments. Following UK devolution in September 1997 centrally managed funds are distributed through separate organisations in England, Scotland, Wales and Northern Ireland. These are shown in Table 2.3.

Table 2.3: Higher Education Funding Councils

Country	Funding Agency
England	Higher Education Funding Council for England (HEFCE)
Scotland	Scottish Higher Education Funding Council (SFC)
Wales	Higher Education Funding Council for Wales (HEFW)
Northern Ireland	Northern Ireland Executive : Department for Higher and Further Education, Training and Employment (DFHETE)

In order to apportion funds appropriately, Higher Education funding councils require data on the size, shape and performance of institutions. These data are acquired through the Higher Education Statistics Agency (HESA), which is the "central source for collection and dissemination of statistics about publicly funded UK Higher Education" (HESA, 2006). Each Higher Education institution in the UK that receives public funding is required to submit an annual "HESA Return", which are datasets of a standard format detailing those students within the institution. Various details are collected: however the majority of the undergraduate student record is derived from the UCAS data supplied at the end of the application cycle. Institutions are encouraged to maintain and update these data as they can have bearing on those funds available for the following academic year. For example, the calculation for widening participation funding for young participants (aged <21) is based on students being grouped into participation rate bands aggregated into the 1991 ward boundary in which they live (HEFCE, 2005). Therefore, if an institution has erroneous or missing postcode data in its applicant records it may be misappropriated or lose funding because of geocoding errors when converting these postcodes into spatial locations. Therefore the two key sets of data which exist for the Higher Education sector are the UCAS and the HESA datasets, the former of which is specifically associated with undergraduate admissions. HESA and

UCAS have made data available for this study and the variables that are of interest will be discussed later in the chapter.

2.3 What is Higher Education?

Section 2.2.2 discusses how data on Higher Education is created and flows between organising bodies and stakeholders. However, before further discussion on other educational data sources it is essential to define what is meant by "Higher Education" and the types of institution that are included in Higher Education datasets. Higher Education is not necessarily university education and the title an institution holds does not determine its classification as either Higher or Further Education (now the Learning and Skills Sector) because the use of the term "University" is legislated differently depending on the age of an institution. Older (pre-1992) Higher Education Institutions operate under a Royal Charter, whereas newer (post- 1992) ones operate under an Instrument of Government and Articles of Government. Both are now managed by a part of government called the Privy Council who are responsible for changes to institutions' constitutions, and also the use of the name 'University' and 'University College' within their title.

As discussed in Section 2.2.2 UCAS processes UK and overseas applications for Higher National Diplomas, Degree and Foundation Degree qualifications for its member institutions. These institutions could be broadly classified as Higher Education, with a number of exclusions including the Open University and Birkbeck College, both of whom operate distance learning and part time courses, and the University of Buckingham as this institution is privately funded. Higher Education Institutions, as defined by UCAS can be "University" or "University Sector Colleges"; however there is no standard definition of either group. University Sector Colleges are often not universities by title, but do run degree or higher-level qualification in conjunction with a university. Conversely, there are also universities that offer FE qualifications alongside degrees. This confusion is further highlighted in the National Qualification Framework (see Table 2.4). In this framework qualifications of similar attainment are grouped into a series of 6 levels, with level 3 the normal progressive step into Higher Education. Therefore, a further definition of an Higher Education institution could be one that supplies level four courses or above. However, this is often problematic, as many Further Education colleges offer vocational qualifications up to levels four or five, but may not offer general qualifications beyond level three.

Table 2.4: The National Qualifications Framework (Source: UCAS)

Levels	General Qualifications	Vocation Related Qualifications	Occupational Qualifications
Higher Level 5	Higher Level Qualifications		NVQ Level 5
Higher Level 4	Degree	Foundation Degree, HND	NVQ Level 4
Advanced Level 3	GCE A Level / AS	Vocational A-Level (Advanced GNVQ)	NVQ Level 3
Intermediate Level 2	GCSE Grades A*-C	Intermediate GNVQ	NVQ Level 2
Foundation Level 1	GCSE Grades D-G	Foundation GNVQ	NVQ Level 1
Entry Level	Certificate of Educational Achievement		

UCAS mainly manages level four courses (HND and Degree), and as such, many courses at University College Sector Institutions are not administered through the UCAS system, with these courses being classified as Further rather than Higher Education.

2.3.1 Classifying Institutional Groups

Section 2.3 showed that there can be duplication between sector datasets and that the definition of specific education sectors can be convoluted. Section 2.1 described the historical context of Higher Education sector growth as measured by admissions through UCAS and its predecessors. Two key events, the amalgamation of Universities and Colleges Central Admissions (UCCA) and Polytechnic Colleges Admissions Service (PCAS), and then UCAS and Arts and Design Admissions Registry (ADAR) resulted in the analytical agglomeration of large constituent groups of very distinctive institutions, the Universities, the Polytechnics and the Arts Colleges. However, between these main divisions there are a diverse set of policy initiatives which resulted in the establishment of a finer taxonomy of institutions. Through a joint project between the University of Oxford and UCAS the taxonomy in Table 2.5 was created.

Table 2.5: The Oxford University Classification of Institutions (Boliver, 2005)

Category	Examples
Ancient Universities	The University of Cambridge, The University of Cambridge.
Old Universities (up to c. 1900)	University of Durham, The University of Birmingham, Cardiff University.
Old Universities (c. post 1900)	The University of Leicester, The University of Hull, The University of Southampton.
Robbins "New Universities"	The University of York, The University of Essex, The University of Sussex.
Robbins "Technological Universities"	Loughborough University, The University of Salford, Brunel University, The University of Bath.
Post 1992 Universities (ex Polytechnic)	The University of Plymouth, Oxford Brooks University, Middlesex University.
Post 1992 Universities (ex HE College)	Cranfield University
HE Colleges and University Colleges – Generalist	Bath Spa University College, Southampton Institute, Newman College, Chester College of Higher Education.
Institutions for Medical Training	The Royal Veterinary College, St George's Hospital Medical School.
HE Colleges – Art / Design / Drama Specialist	The London Institute, Norwich School of Art and Design, Wimbledon School of Art.
HE Colleges – Other Specialist	Greenwich School of Management, Scottish Agricultural College.
FE Colleges	Barnsley College, Bradford College, City of Bristol College.
Private Institutions	SAE Institute Regents Business School, University of Buckingham.

Although these grouping provide a useful framework to understand the history of how different types of university were established, they are not agglomerations used in policy making, nor lobbying. Groupings of institutions which have formed outside of historical coincidence, and with a specific remits include:

- The Russell Group
- The 1994 Group
- N8 Group

The Russell Group[2] is the largest aggregation of research intensive institutions, set up in 1994 as lobbying group to promote their interests to government and associated bodies. This group accounts for 65% of the total UK university research grant income, amounting to around £1.8 billion in 2006. Not all "research intensive institutions" are part of the Russell Group and in response to this the 1994 Group[3] was set up between those who were excluded. The main difference between the two groups is that the Russell Group institutions tend to have Medical Schools and a scientific focus. The 1994 Group took on a similar remit to the Russell Group of lobbying government but are less influential because of their smaller sizes and research incomes. The N8 Group are a partnership of eight research intensive

[2] www.russellgroup.ac.uk
[3] www.1994group.ac.uk

institutions including Durham, Lancaster, Liverpool, Leeds, Manchester, Newcastle, Sheffield and York. The initiative was created as part of a joint project between three Regional Development Agencies in the North of England. Unlike the Russell Group that is concerned with lobbying, this group focus on a series of joint research agendas. A final point on the classification of institutions is that institutions can change their names or indeed composition, and as such when comparing temporally these changes much be understood. The definition of individual institution can change over time such as the changing of names or the amalgamation or de-merger of institutions. One recent example was the 2004 merger of the Victorian University of Manchester with the University of Manchester Institute of Science and Technology. Together these institutions became the University of Manchester, and obviously any statistics created to examine pre and post 2004 should reflect these changes.

2.4 Prior Qualification Data

As outlined in Section 2.2.1, prior to Higher Education applicants obtain entry qualifications through the school and Further Education sectors. Compulsory education currently runs up to the age of 16 and culminates in GCSE exams or their equivalent level qualifications, after which the student can study in post compulsory education or enter full time employment. Post compulsory education is usually required to gain entry qualifications suitable for Higher Education study. Compulsory education is divided into a series of progressive Key Stages which represent standards which are set out in a National Curriculum that students should have met by these points in their education. The Key Stages are as follows:

- Nursery School
  - Key Stage 0 (3-5 years old).
- Primary School
  - Key Stage 1 - Years 1 to 2 (5-7 years old)

  - Key Stage 2 - Years 3 to 6 (7-11 years old)

- Secondary School

  - Key Stage 3 - Years 7 to 9 (11-14 years old)

  - Key Stage 4 - Years 10 to 11 (14-16 years old)

- Sixth Form School / College / Further Education College

  - Key Stage 5 - Years 12 to 13 (16-18 years old)

At the end of Key Stage 2, 3, 4 and 5 students sit a series of tests which in the case of Key Stage 4 and 5 lead to an individual qualification, however in all Key Stages these results are used to construct the annual DCSF schools attainment tables[4].

There are a series of factors which make data collection, analysis and dissemination for the UK school sector complex. Firstly, since devolution there are different organising bodies for each of the countries which make up the UK. In England this is the Department for Children Schools and Families (DCSF)[5], in Scotland it is the Scottish Executive (SE)[6], in Wales it is the Department for Education Lifelong Learning and Skills (DELLS)[7], and in Northern Ireland it is Department of Education[8] (DoE). Some curriculum differences occur between the countries. In all countries other than Scotland, Key Stage 4 culminates in GCSE or their equivalent qualification. However, in Scotland, Scottish Standard level qualifications are studied. Similarly, in Scotland at Key Stage 5, Scottish Higher levels are studied, where as in the rest of the UK A-Levels or their equivalents are taught. A further complication with school data is that independent schools do not have to submit their pupil demographic data to those government organising bodies discussed earlier, thus demographic data for the independent sector is not publically

[4] http://www.dfes.gov.uk/performancetables/
[5] http://www.dfes.gov.uk
[6] http://www.scotland.gov.uk/Topics/Education
[7] http://new.wales.gov.uk/topics/educationandskills/?lang=en
[8] http://www.deni.gov.uk/

available. In post compulsory education at Key Stage 5, students progressing from Key Stage 4 do not necessarily study in schools. Some students opt to study in Further Education colleges who have a different organising body, the Learning and Skills Council (LSC).

2.4.1 Sources of Prior Qualification Data

In Chapter 9 both Key Stage 4 and 5 data are examined for England only. Maintained schools and colleges receiving public funding have had a statutory duty to supply data to the DCSF on an annual cycle since 2002 (Jones and Elias, 2006). These data are stored at the DCSF in the National pupil database (NPD) across a number of datasets (See Figure 2.10). The linking field which can be used to join demographic data to attainment data at the level of the individual is referred to as the Pupil ID. The records of individuals can also be linked to a unique school identifier called a unique pupil number (UPN) to create a series of analysis examining both demographics and attainment within schools. It is these data which are used to calculate school performance tables in England.

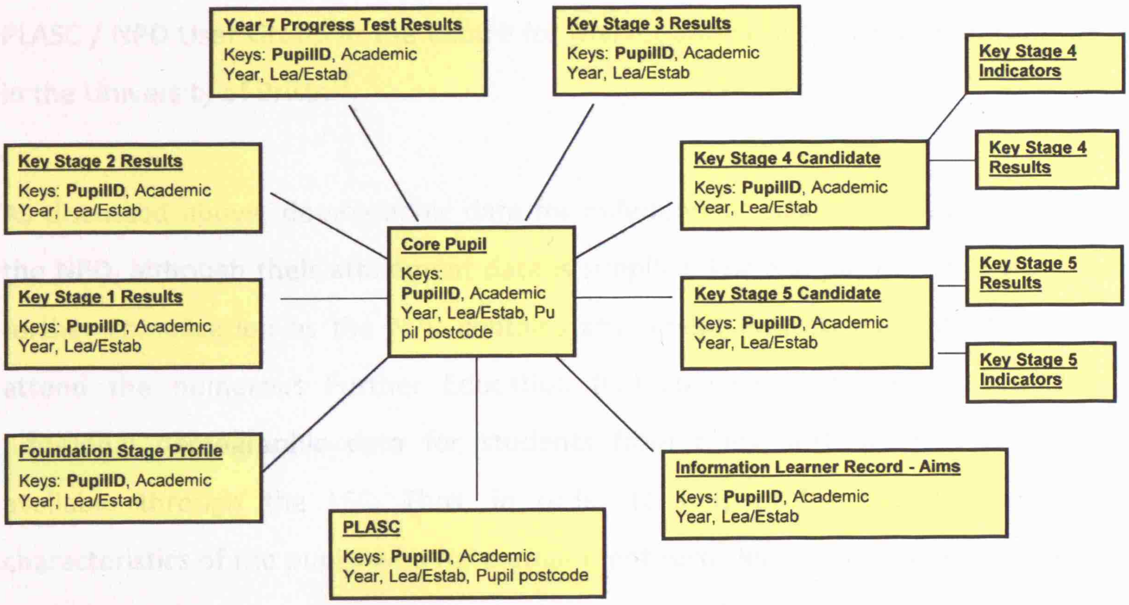


Figure 2.10: The National Pupil Database (Barker, 2006)

The attainment files supplied by the DCSF for research purposes contained attainment results for all students, but personal / demographic data (e.g. unit postcodes) only for those who are included in the Pupil Level Annual Schools Census

(PLASC). PLASC is a survey of all students in publicly funded state schools and captures a range of demographic data including:

- Forename and Surname
- Postcode
- Ethnicity
- Free School Meals Eligibility
- Disability Status
- Language of Origin

For access to data at the unit postcode level, special access requirements were negotiated as these are classified as sensitive data. These data were supplied by the PLASC / NPD User Group at the Centre for Market and Public Organisation (CMPO) in the University of Bristol[9].

As discussed above, demographic data for independent schools is excluded from the NPD, although their attainment data is supplied. For analysis of KS5 there is a further complication as the NPD contains attainment data for students for who attend the numerous Further Education (FE) colleges managed by the LSC. Additional demographic data for students from these institutions was made available through the LSC. Thus, in order to fully profile the demographic characteristics of the pupils who have attainment recorded in the NPD at Key Stage 5, multiple data sources were required. Access to demographic data for those independent schools discussed in Chapter 9 was provided by the Independent Schools Council (ISC)[10] who granted access to a selection of the Independent Schools Council Pupil Database (ISCPD). All those data required to profile both

[9] http://www.bris.ac.uk/Depts/CMPO/PLUG/
[10] http://www.isc.co.uk/

demographic and attainment characteristics of schools in KS5 are shown in Figure 2.11. In order to measure demographic characteristics of attainment at the pupil level, a match is required between all the demographic and attainment datasets, using an identifier unique to individuals. Unfortunately this does not exist and as such the Key Stage 5 analysis in Chapter 9 focuses on the school level analysis only. Furthermore, the DCSF, LSC and ISC all use a different coding scheme to identify individual educational establishments and as such in any large / national scale analysis a lookup table would have to be created, possibly using fuzzy matching on the address and school / college names.

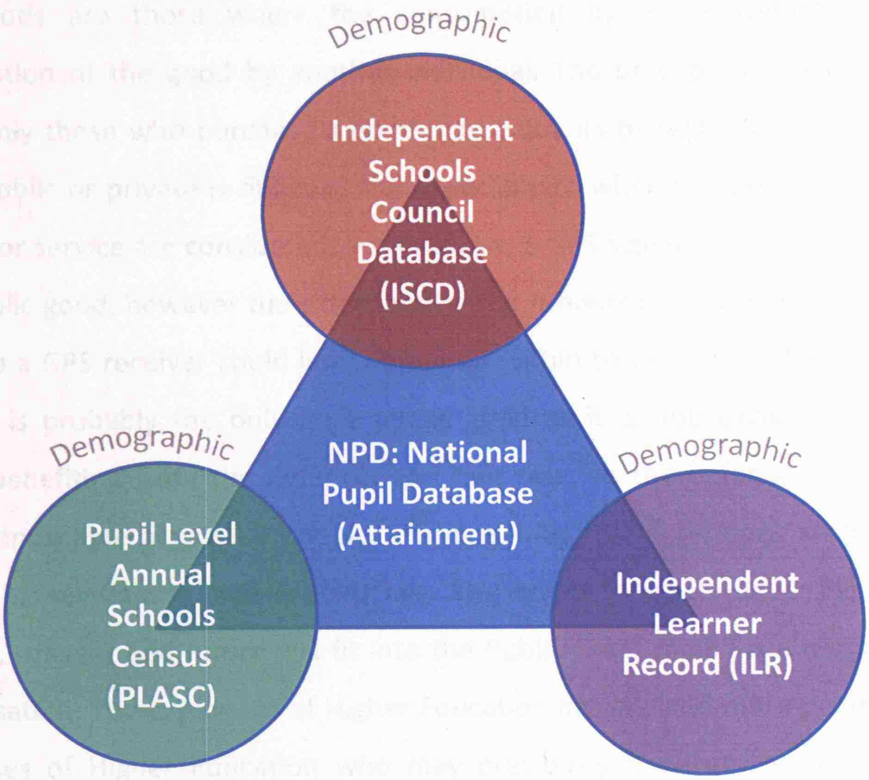


Figure 2.11: Key Stage 5 Data Sources

2.5 Is Higher Education a Public or Private Good?

Section 2.1 discussed how participation in Higher Education is growing however access remains stratified by socioeconomic group. The importance of access equality depends on the extent that Higher Education goods and services are classified as public or private by their consumption and benefit limit. Public and Private goods fall within the framework outlined in Table 2.6 (Peston: 1972:13).

Table 2.6: Public and Private Goods Framework

	Excludable	Non Excludable
Rival	Private Good	Common Pool Resource
Non Rival	Toll Good	Public Good

Rival goods are those where the consumption by an individual limits the consumption of the good by another individual. The principle of excludability is where only those who purchase a good may enjoy its benefit. Defining a good as purely public or private is difficult, and especially so when the externalities of the product or service are considered. For example, a GPS signal could be considered a pure public good, however the externalities that it benefits only those directly with access to a GPS receiver could limit definition within this category. The Ministry of Defence is probably the only pure public good as it is non excludable, with all people benefiting from the defence, and non rival, as the defence is reasonably equally applicable across all people, perhaps with few exceptions such as extra defence for key figures such as politicians. The extent to which Higher Education in the past, present and future will fit into the Public Good model is a more difficult categorisation. The expansion of Higher Education has allowed more people access to courses of Higher Education who may previously have not attended and is pushing the model towards non excludability, however the extent to which Higher Education is non-rival is debatable, and in 2007 this depends on whether one is applying for a place at a selecting or recruiting institution. Selecting institutions are those that have capacity short of their demand and as such become more selective of those students who are made offers. A recruiting institution is one that can usually accommodate all applicants with the minimal requirements to undertake a chosen course, and as such may for example be more pro-active with marketing to fill places. In reality, the extent that institutions fit within this classification will be graduated; however for the purposes of explaining of relative institutional behaviour it provides a useful dichotomy. The equality of access issues discussed earlier undermines the non-excludability principle, as there are certain groups within society who are underrepresented in Higher Education. Therefore Higher Education does not have the characteristics of a standard public good. However, Creedy (1994) discusses that public investment is justified as Higher Education is at risk of market failure because of the long term horizon of returns to the individual, and that consumption will not be at the socially optimal level without subsidy. The prevailing argument to support the public finance of Higher Education is that it generates externalities that are not directly apportioned to the individual, and as such are beneficial to wider society.

The 2003 White Paper on Higher Education announced that the Higher Education funding gap will be rectified by the introduction of quasi market forces which is contentiously at odds with the government's past image of a public good. Furthermore, the admission by the then education minister Charles Clarke that not all institutions are the same in terms of the opportunities and life chances that result from courses of study further adds to this debate (Collins, 2003). This process has been referred as the 'marketisation' of Higher Education, however, there is significant evidence to suggest that the public already think of Higher Education in market terms, even with current capped fee prices. League tables are published each year by major newspapers to assess the ranking of particular institutions or courses, and the Research Assessment Exercise grades departmental research output against global benchmarks. Furthermore, the reputation of certain institutions affects the perceived value of courses of Higher Education. For example, a degree from the University of Oxford and Cambridge University will be seen by the majority as more valuable than one from an ex-Polytechnic College.

The true individual financial return from a degree is impossible to empiricise absolutely, however in 1992 the DfES estimated that graduates would earn over their lifetime £400,000 more than non-graduates (Hodge, 2002a), and the 2003 Higher Education White Paper (DfES, 2003) discusses that this averages around 50%

more than non-graduates. These estimated financial benefits have been used in supporting arguments for the introduction of higher variable tuition fees, however they have also been widely criticised. Aston and Bekhradnia (2003) discuss that the 50% earning figure relies on speculative predictions that are impossible to measure accurately. Blasko (2002) adds that the socio-economic background of graduates also influences their relative success in the labour market, and as such the complexity of cultural and social capital would need to be factored into any representative economic degree premium calculation. Therefore, although quantifiable financial benefit of Higher Education participation is debatable, it seems clear that the skills it provides, and the life chances that result are all of great benefit to the individual. The economic or societal benefit described in the 2003 White Paper "suggests that there is compelling evidence that education increases productivity, and moreover that Higher Education is the most important phase of education for economic growth in developed countries, with increases in Higher Education found to be positively and significantly related to per capita income growth" (DfES, 2003:58). This is an interesting point as individual total lifetime productivity is diminished by the three or more years it takes to gain a degree, and as such this suggests that the extra productivity Higher Education allows to develop is greater than the duration of the course. However, one criticism is that this does not make any reference to the relative productivity benefit for particular courses, assuming the benefit to society is the same across all subject classifications.

In absence of a reduction in real terms, or through re-labelling of institutions categorised as universities, the partial introduction of market forces by the Higher Education White Paper provides an improved but not perfect financial structure which will allow institutions the autonomy to invest in more suitable widening participation strategies. Accepting that institutions have different needs is a step forward, and institution level differentiation within the sector is fundamental to this model being success. The introduction of market based funding will encourage institutions to embrace their differences through evaluating strengths and weaknesses relative to competitors in order to employ strategies to win market share.

It is argued by some institutions that the current £3000 fee limit does not go far enough to rectify the deficits in their finances. Sir Howard Davies, director of the London School of Economic stated that the £3000 maximum fee will only halve the deficit of this 'loss making business' (Economist, 2004). The majority of selecting institutions charge the full £3000, and because of their over subscription, the 'market' could probably sustain a far higher price without significant loss in applications. Many of the selecting institutions have announced financial participation incentives which have been introduced with these top up fees. Middlesex University has introduced £1000 bursaries to students who gain places with at least 3 B grades at A-Level or equivalent (Macleod, 2003). Furthermore, Royal Holloway offer bursaries for a single year's postgraduate tuition after successful completion of their undergraduate studies (BBC, 2004).This supports the marketisation principle in that some institutions are starting to operate more market led recruitment strategies where well qualified students can trade good A Levels for cheaper admission.

Proponents argue that institutional ability to charge variable fees, combined with decentralised funding sources will create, in a more traditional sense, a market led Higher Education system that will allow Higher Education institutions to better determine their future (Smithers, 2002). However, there are those who believe that universities are not yet ready to adopt this model, and that our current market-state hybrid system is the worst of both worlds. Scott (2002) identifies several key problems with the introduction of variable fees, discussing that traditional universities such as those members of the Russell Group may be inclined to push up fees not to satisfy market conditions but to protect their own university brand. These universities may not wish to be seen to charge bargain prices as it may reflect on the perceived quality of their products or courses. It is further argued that a positive effect of this could be for universities that currently are at the bottom end of the market as they may be able to undercut the market leaders possibly through offering reduced fees or attendance incentives, therefore creating new market share. The middle market, made up of the bottom end of the old institutions and the top of the new may be in constant flux. Scott (2002) proposes that some institutions will provide niche courses while others will combine to reinforce their brand. As Phoenix (2003) contends, it is certain "the traditional bilateral relationship between Higher Education and the state is rapidly becoming a multilateral relationship between Higher Education and various external funding bodies". These external funding bodies refer to industry, overseas recruitment and student fees, all of which are adopted in differing mixes to form our current hybrid state-market controlled Higher Education system.

Sir Howard Newby, former Chief Executive of the Higher Education Funding Council for England recently announced:

> ***"We worry that some institutions might get this wrong [referring to fee price]. They think they can sustain a £3000 fee across the board, when they will actually find they can't"***

HEFCE (2004:1)

This decree came before the more recent announcement that some institutions have been considering a strike price to attract targeted levels of students with given levels of attainments and probably also consistent with intended institutional profiles. Sanders (2004) discussed that at this time both Leeds Metropolitan and Bradford Universities were putting together proposals for fees as low as £2000 across the board.

Lessons from an international perspective teach us that it is very difficult to predict the actual impact fee levels will have in terms of participation and access. In Australia there has been an increase in participation since fees were introduced in 1989, however opponents do argue that access has contracted during this period. Ramesh (2004) also discusses how graduates from the Indian Institute of

Management were dissatisfied with proposals to cut student fees in an attempt to widen access. One student remarked 'the government's logic is completely wrong. The fees are high because the facilities here and the Professors are world class and someone has to pay for them'.

Market led activities will demand greater intelligence about competitor institutions, their customers' characteristics and, crucially, where they live and are educated. Although a growth area in public policy, there has been relatively little research to date on the specific exploitation of geodemographic techniques, data and tools within the Higher Education context. Two notable exceptions are Tonks (1999) and Tonks and Farr (1995), who examine the applicability of the language and tools of marketing within a Higher Education context. Geodemographic analysis can and will play a key role for institutions to gather essential profiling information, applying tools and techniques more accustomed with those utilised by the private sector to target products and services at specific market segments.

2.5.1 "Quality" and the Role of Rankings

The progressive "marketisation" of Higher Education introduced in Section 2.5 has encouraged a number of developments that would be associated with consumer orientated markets. There are numerous data sources in the public domain which allow applicants to make informed decisions about which institutions or courses are considered to be "high quality". The definition of "research quality" is assessed by the Research Assessment Exercise (RAE)[11] and the "teaching quality" assessment by the Quality Assurance Agency (QAA)[12]. The quantification of these institutional quality measures provide students with a wealth of information about the institutions that would provide them with the "best" degrees for their chosen subject. In addition to this information, HESA publish annual data on various aspects of Higher Education institutions, such as staff student ratios and research income. These data could also be used to support applicant decision making. The dissemination of both quality assessments and HESA data are not in an easily

[11] Details on current and previous RAE can be found at: http://www.hero.ac.uk/rae/
[12] Details on teaching assessment can be found at http://www.qaa.ac.uk/

digestible format for potential applicants, nor are they available from a single location and as such have given rise to a series of consumer orientated private sector metrics of Higher Education performance. These metrics take the form of annual guides containing institution and subject rankings using a range of data.

The guides include:

- Good University Guide (The Times)[13]
- The Sunday Times University Guide[14]
- Guardian University Guide[15]

The ranking methodology and data used are similar across the two Times guides, but differ slightly in the Guardian. The Guardian classification does not include any data from student satisfaction surveys, for example. Ranking of data is designed to give applicants an informed choice of Higher Education institution, but because the rankings are created from weighted input variables the actual order in which institutions appear in these lists can be manipulated greatly through adjustment of these weights. Using the data which were input into the 2003 ranking, an application was built in which different weighting scenarios can be tested (See Figure 2.12).

[13] The guide is available from www.thegooduniversityguide.org.uk
[14] Available from www.timesonline.co.uk/tol/life_and_style/education/sunday_times_university_guide
[15] Available from http://education.guardian.co.uk/universityguide2006

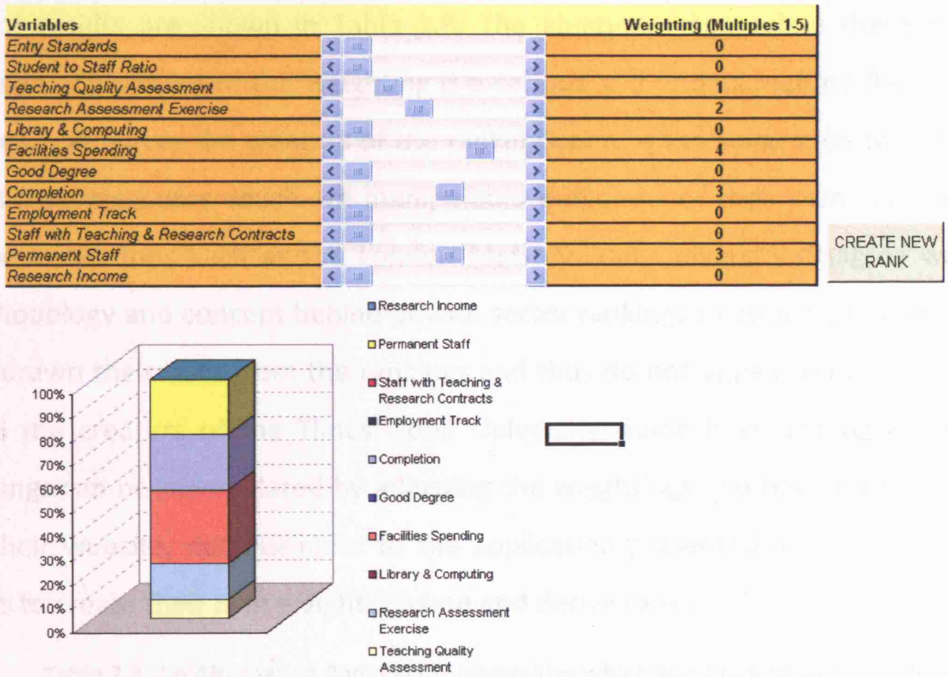


Figure 2.12: An Example University Ranking Application

This application adjusts the weights for the input variables, calculates the new weighted variable scores and then creates an overall score for each institution. A Visual Basic macro then creates a new sorted list based on these data. For example, Table 2.7 demonstrates a ranking which was created where RAE and Entry Standards were weighted 5 times higher than all other variables.

Table 2.7: An Illustrative Example of University Ranking where RAE Ratings and Entry Standards were Weighted Highly

Rank	Institution
1	Imperial
2	Oxford
3	Cambridge
3	UCL
5	Sussex
6	UWIC, Cardiff
7	LSE
8	King's
9	Bristol
10	Edinburgh

Using the same data, however this time weighting student to staff ratio and the Teaching Quality Assessment five times higher than the other variables a different set of results are shown in Table 2.8. The ability to manipulate these rankings through adjustment of the weighting is a serious and undocumented flaw in these guides, and leaves the creators of the ranking open to the suggestion that they are subjective measures which are manipulated outcomes of their own predilections. Some institutions such as the London Metropolitan University disagree with the methodology and concept behind private sector rankings so strongly that they have withdrawn their data from the rankings and thus do not appear on the lists. Since 2006 the creators of the Times Good University Guide have acknowledged that rankings can be manipulated by adjusting the weightings and have provided a tool on their website, not dissimilar to the application presented above which allows users to create their own weight scheme and derive ranks.

Table 2.8: An Alternative Ranking of Universities which Bias Student to Staff Ratio and the Teaching Quality Assessment

Rank	Institution
1	UWIC, Cardiff
2	Imperial
3	Sussex
4	Oxford
5	Salford
6	Cambridge
7	UCL
8	Bath
9	King's
10	Bristol

2.6 Causes of and Solutions to Inequality

Within a context of growing marketisation, access inequality has remained. Reid (1998) discusses that there are two interpretations of inequality in Higher Education: first, that there is bias in the university selection process; and second, social class has an inhibitor effect on the perceived availability or benefits of Higher Education. The first of these interpretations was publicly highlighted in 2001 with the case of Laura Spence. Her rejection by the University of Oxford on the basis that she "did not show potential" created a media circus that even involved the then Chancellor of the Exchequer (and now Prime Minister) who declared it "an absolute scandal". The second of these interpretations relates to how middle class parents 'invest all kinds of effort, including significant material resources in developing social capital' (Walker, 2003:172), creating environments where socialisation processes can occur, and creating advantage or disadvantage under certain situations (Bourdieu & Passeron, 1977). Social capital may be defined as the advantage conferred over non-group members through interaction within a network of individuals, who often share similar beliefs or values, and that ultimately lead to greater group-wide economic or social gain. This is not dissimilar to the concept of cultural capital and the two concepts have often been interlinked. Social or cultural capital confers an individual benefit or disadvantage under certain social conditions, such as feeling 'comfortable' or enabling interaction with peers within a particular Higher Education institution. Interrelated with ideas of social and cultural capital is the method by which an individual experiences and perceives space. This 'space' could refer to an individual perception of whether Higher Education is accessible or restricted. The study of these perceptions is referred to as cognitive mapping, and although this thesis will not investigate this topic in depth it is mentioned as an attitude forming framework from which behaviours are measured. Kitchin and Blades (2002:7) discuss that 'cognitive maps provide insights into the relationship between people's environmental representation and their behaviour in the environment'. Cadwallader (1976) quoted in Kitchen and Blades (2002) discuss that at least three types of spatial decision are influenced by an individual's cognitive map. These are the decision to stay or go, where to go and finally which route to take. The analogue of this concept to individual decision-making in Higher Education is high, and the extent to which different cognitive maps develop as a result of societal interactions, be this in terms of social class or geodemographic area effects, will be measured in terms of the behaviours recorded in the Higher Education datasets. Knowledge of these behaviours will provide insight into how the cognitive maps of different typologies allow for variable decision-making processes at the area level. Indeed, Krech *et al* (1962:20) discuss that the cognitive map is "a partial personal construct in which certain objects, selected out by the individual for a major role, are perceived in an individual manner".

The extent to which one can intervene directly in these access patterns is the realm of social engineering. The introduction of class stratification limits and typology targets are viewed as unethical in the United Kingdom, and any effort to introduce such measures are met with strong resistance by both the institutions and bodies representing the interests of individuals deemed to be disadvantaged by the proposed targets. These issues discussed in general terms by the Economist (2004):

> ***"Micromanaging university admissions, as the British government has been trying to do on grounds of class, with targets, quotas, fines and strictures, risks the same consequences as similar American experiments based on racial preference. It humiliates the talented but disadvantaged, whose success is then devalued; it infuriates the talented who are not deemed underprivileged enough and who feel their merits ignored, and it makes universities do a job they are bound to be bad at."***

The Economist (2004: 23).

Furthermore, a specific example of resistance to perceived social engineering occurred when UCAS announced that from 2008 the application form would include a question on whether applicant's parents had entered Higher Education. Pat Langham, the president of the Girls School Association discussed that "favouring candidates whose parents didn't go into Higher Education is artificial and amounts to social engineering" (Lightfoot, 2007).

However, where an institution is unaware of relative performance indicators, and also ways in which various conceptions of capital influence an applicant's prior development and attainment, discrimination can occur as decisions are based on partial information. Perhaps a high achieving applicant to the University of Oxford or Cambridge from a poor performing comprehensive school would not have access to the same literature as an applicant from Eton, therefore the expectation of knowledge of such texts in interview questioning by admissions staff could be perceived as discriminatory. However, this does pose a problem. If a wide literary knowledge is an essential requirement to meet the rigorous demands of a particular course, an admissions tutor without background information may assume a particular candidate is unsuitable, and thus not offer a place. If the admissions tutor had access to contextual information about the applicant background, this inadvertent discrimination could be avoided, perhaps through supply of introductory readings before commencement of the course.

However, if this information were known, and the desirable quantities of individual groups specified were specified by policy, then these activities could be considered methods of social engineering – that is attempting to compensate for the failures of state education and social policy by crude adjustments to institutional acceptance profiles. As such, when examining those influences on applications and acceptances within Higher Education, it is important to seek to accommodate such considerations. A key aim should be to effectively extend participation to those segments in society whose participation is currently disadvantaged by internal and external social or cultural values. However, incorporating these ideas into what is historically a tiered applications system will not be without controversy. Those schools that have always sent their pupils to particular universities will resist measures that would result in these patterns changing. Pauline Davis of the Girls Schools Association suggests "it will be difficult, if not impossible, for many of our students to demonstrate exceptional performance in context since the pupils who attend our schools achieve such high standards" (BBC, 2002). This echoes the sentiments of the Headmasters and Headmistresses Conference[16] that represents the views of 250 leading public schools. It produced an investigation in 2002 that showed how in the worst case 80% of their pupils were being rejected without interview on certain courses in Russell Group Institutions, claiming that this was a result of positive discrimination policies in these universities (Guardian, 2002). Grimson and Dobson (2002) agree, arguing that numerous universities have introduced schemes to increase the total number of state school students without increasing the total number of students, therefore squeezing applications from independent schools. However, these criticisms ignore evidence to suggest that independent school pupils gain lower degree scores than their state educated equivalents. Allison (2002) discusses that an eight year study of every graduate in the UK revealed independent school students had an 8% lower chance of obtaining

[16] The Headmasters' and Headmistresses' Conference (HMC) represents the Heads of some 250 leading independent schools in the United Kingdom and the Republic of Ireland. (http://www.hmc.org.uk/)

a first or upper second class honours degree in comparison to a state school educated pupil with the same A-Level results.

Professor Steven Schwartz published the final version of the Admissions to Higher Education Review in September 2004, which has investigated in depth issues of equality in Higher Education admissions. In it, he suggests "school type tends to distort the predictive or signalling ability of prior attainment" and that "school performance may also affect the predictive ability of prior attainment" (AHERG, 2004:45). This does raise an interesting question to whether school type is a direct or indirect indicator of social capital formation. If school type is considered a direct indicator of social capital formation then attendance will lead to a greater advantage when applying for Higher Education, given that this is a usual and supported course of progression for individuals within these groups. This may occur by being offered better admissions advice when applying, or, could be a self-reinforcing phenomenon with each successive generation attending Higher Education, assuming in this social capital model that the perceived socioeconomic benefits outweigh the costs. If school attended is used as an indirect indicator of social capital formation it may be that the applicant would have made attempts to enter Higher Education independent of whether they attended a particular school type. The school may only reinforce the decision or confer better chances of application success rates. Therefore an underlying focus of the investigation throughout this thesis is on the nuances of these measures and attempts to generalise their relative importance.

There has been a shift in focus by the media from purely identifying those institutions who are deemed 'fair' and 'unfair' in their admission practices, to issues caused by the extension of access into underrepresented groups. Hall (2001) discusses that cost of non completion of programmes of Higher Education in 1995-96 was around £91.5m. In 2002 this figure was re-estimated for the current year at £150m, and the causation attributed to recruiting poorly prepared students (Clare, 2002). The study showed that students from lower socioeconomic groups had lower

A-Level scores, were applying through clearing and were more likely to drop out of courses. Thomas (2002) discusses that the main factors influencing retention rates include:

- Unprepared for Higher Education
- Prior Academic Experiences
- Institutional Expectations and Commitment
- Academic and Social Match
- Finance
- Family Support
- University Support

The first two points relate how prior academic experience in the type and nature of qualifications create different support requirements between groups entering Higher Education. The third influence is the extent to which an individual has their expectations met by an institution. The fourth influence on retention is how academically and socially an individual feels they are part of an institution. Finance also affects retention, and in particular the extent to which a student feels they are undergoing financial hardship. Family support and university support are related to the extent to which an individual feels they are emotionally and academically supported in their choice of institution and course. Retention studies have investigated within the context of individuals and the social class to which they belong. As individuals live and interact within areas, there is a plausible assumption to be made that these effects should also be of influence to retention.

2.6.1 Pre Higher Education Performance

The measurement and interpretation of access rates in Higher Education are often made in ignorance of how attainment stratification is embedded far earlier in the applicants educational histories. However, in the 2003 Higher Education White Paper (DfES, 2003:68) the DfES has accepted that 'the single most important cause of the social class division in Higher Education participation is differential attainment in schools and colleges'. Furthermore, Leathwood and Hutchings (2003:137) discuss that prior attainment fits within class profiles and that 'working class pupils continue to do less well educationally than their middle class peers' progressively throughout their educational careers. They set out possible reasons for the attainment gap as:

- Poverty
- Family Expectations
- Classed Assumptions about Ability
- Gendered Assumptions about Ability
- Race Assumptions about Ability
- Cultural Capital
- Parental Involvement in Schooling
- Cultures and Practices of the Educational Institutions

However, these reasons negate to mention neighbourhood effects, which could be very influential as most of the processes above are imbedded in different geographies across a broad number of scales.

Selecting institutions are faced with a further achievement problem when selecting the most suitably qualified candidates, in that the numbers of high achievers are growing. Top institutions such as Cambridge University have introduced extra measures to aid candidate selection such as entry tests (Ward, 2003), and calls for candidate A-Level module results (Hackett, 2003). Nationally UCAS have replaced the A-Level points system with a tariff score, which incorporates a plethora of other qualifications under the premise that offers will be given in a tariff score, not A-Level points. To illustrate an extreme example a candidate can have 360 tariff points by possessing 3 A grades at A-Level, or, could have an A in the practical and theory CACHE Diploma in Child Care and Education. For a selecting university this would pose a problem, as the 'equally qualified' candidates may have the tariff requirements for a 360 point course, but in reality would they be suitably qualified? Ryan (2004:13) discusses that 'for almost any course, what a student has done in particular areas is more important that whether they have 120 or 140 points'.

The Tomlinson Review published in October 2004 sets out recommendations that were designed to dramatically alter the pre-HE curriculum. This review was justified in Higher Education terms as follows (DfES, 2004):

- Too few young people continue learning beyond compulsory schooling
- It has become increasingly difficult to differentiate between high achievers
- Too few young people have the right skills
    - Communication
    - ICT
    - Number
    - Research Skills
- Too few vocational qualifications meet the needs of learners, Higher Education and employers
- The system is confusing and unclear

The Tomlinson Review proposed that these problems are solved within the new three-stage diploma framework outlined in Table 2.9.

Table 2.9: An Outline of the Three Stage Diploma Framework (DfES, 2004)

Diploma level National Qualifications	Framework level	Existing national
Advanced	Level 3	Advanced Extension Award; GCE AS and A level; level 3 NVQ; equivalent qualifications
Intermediate	Level 2	GCSE at grades A*-C; intermediate GNVQ; level 2 NVQ; equivalent qualifications
Basic	Level 1	GCSE at grades D-G; foundation GNVQ; level 1 NVQ; equivalent qualifications
Entry	Entry	Entry Level Certificates; other work below level 1

However, despite these propositions the Swartz diploma classification was not implemented and instead the current GCSE and A-Level system were kept. A recent development to tackle the issue of discriminating between the top achieving students is that the Qualifications and Curriculum Authority have announced that A-Levels are to be given an extra grade of A*, where students should attain at least 90%.

2.7 Conclusion

This chapter has introduced some of the literature and arguments that underlie and provide context to the empirical analysis throughout this thesis. Higher Education is in a sustained turbulent period of growth in which it has transformed from a minority to a mass market system, with very rapid growth over the last 50 years. Various initiatives have created numerous organising bodies, each with their own data and collection mechanisms. The integration of these data is currently at best partial; there exists overlap within the data collection mechanisms and for some sectors such as independent schools demographic data are not collated nationally. Funding for Higher Education is presently improving, although there remains a legacy of underfunding which has left many institutions with large deficits. In a mass model of Higher Education, basic marketisation has been actioned through the introduction of variable fees as a method of rectifying these deficits and sustaining an internationally competitive sector. However, this chapter has begun to highlight ways in which broadening access to Higher Education in terms of absolute numbers may not have the effect of extending access to all groups equally.

3

SOCIO-SPATIAL DIFFERENTIATION.

3.1 The Science of Classification and Taxonomy

Longley *et al* (2005:11) discuss how "information systems helps us to manage what we know by making it easy to organise and store, access, retrieve, manipulate and synthesize, and apply knowledge to the solution of problems". Because the world is complex, humans have an intrinsic desire to classify reality and to seek an ordering framework through which information as perceived may be assembled and understood. Bowker, *et al*, (1999:1) discuss that at a basic evolutionary level "to classify is human" because "human physical abilities are limited, so the amount of information provided to us is constrained by our ability to see" (Weinberger, 2007:4). Humans rarely if ever perceive every detail of reality, and as such, create internal cognate models through codifying observations using appropriate levels of detail. These processes are formalised in Psychology through schema theory which describes the construction of those "mental representations which are used during perception and comprehension, and which evolve as a result of these processes" (Anderson, 1977:418). Rosch *et al* (1976: 382) further posit that these categorisations "are not arbitrary but highly determined", that is there are basic and common categories of understanding between all humans. These occur because:

> ***"The world is structured because real world attributes do not occur independently of each other. Creatures with feathers are more likely also to have wings than creatures with fur, and objects with visual appearance of chairs are more likely to have functional sit-on-ableness than objects with the appearance of cats. That is, combinations of attributes of real objects do not occur uniformly. Some pairs, triples, or ntuples are quite probable, appearing in combination sometimes with one, sometimes another attribute; others are rare; others logically cannot or empirically do not occur"***

Rosch *et al* (1976:383)

Rosch *et al* (1976:383) define a *category* as "a number of objects which are considered equivalent", and a *taxonomy* as "a system by which categories are related to one another by class inclusion". As suggested by Schema Theory, taxonomy of categories therefore provides the mental framework through which humans can understand the world. This is exemplified by Davis (1991: 21) who relates a schema of an egg to a series of related categories and taxonomy (See Figure 3.1).

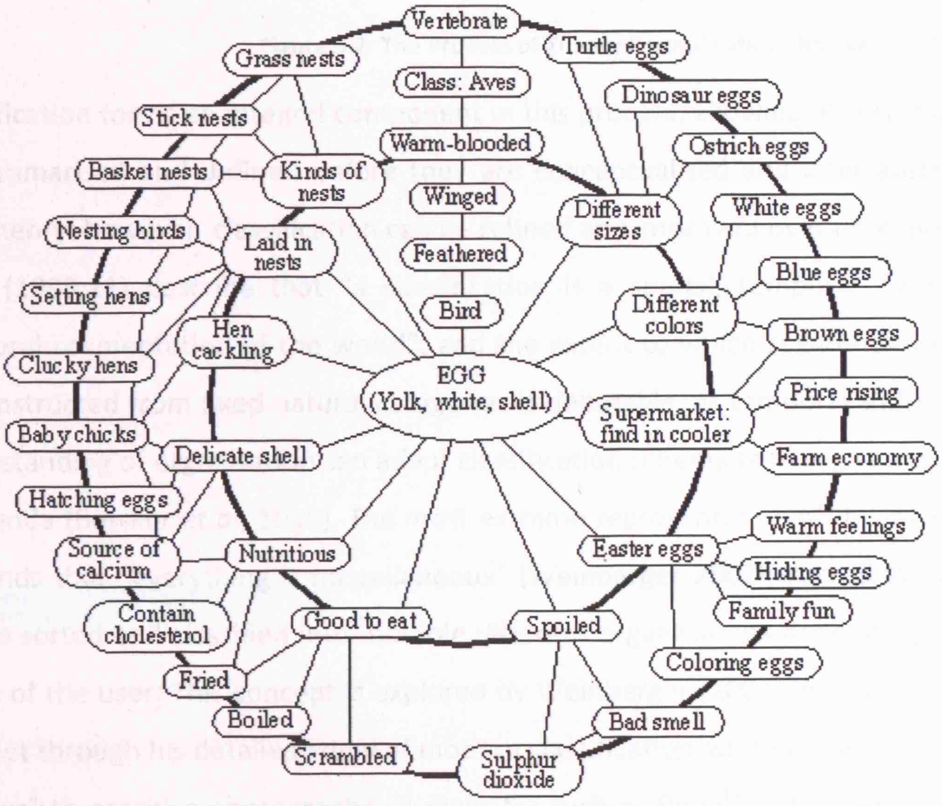


Figure 3.1: An Example Schemata of an Egg (Davis, 1991:21)

These cognate generalisations mirror the competing needs of ideographic and nomothetic science; that is the tension between analyses of the uniqueness of objects versus generalisation of processes. Longley *et al* (2005:14) discuss how the technology of GIS, for example, can be used to "combine the general with the specific" through capturing and implementing general knowledge in software while maintaining a database which represents specific knowledge. Within this context, *appropriate* classification can be considered as a problem solving tool which software may use to organise those tacit attributes on the uniqueness of spaces, in order that processes under observation may be understood through empirical observation. The understanding of unique events and processes has been framed as theory formulation by Bracken (1981:112) who discusses "[a] subject without theory is all fact, because facts, or observations, are then only what is *believed* to be correct". This process of theory formulisation is shown in Figure 3.2.

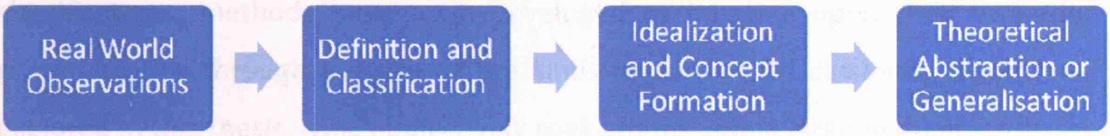


Figure 3.2: The Process of Theory Formulisation (Bracken, 1981: 113)

Classification forms an integral component in this process, allowing observations to be summarised and defined before they are conceptualised and later abstracted into theory; however, classification can be refined and improved over time. Bowker, *et al* (1999:11) describe that "a classification is a spatial, temporal, or spatio-temporal segmentation of the world", and the extent to which segmentations can be constructed from fixed natural categories is debatable, as temporal shifts in our understanding of organisation can adapt classification schema to fit a given purpose or agenda (Bowker *et al*, 1999). The most extreme representation of this definition contends that "everything is miscellaneous" (Weinberger 2007), such that objects can be sorted and classified into multiple different organisations depending on the needs of the user. This concept is explored by Weinberger (2007) in relation to the Internet through his detailed study of modern classification schema, such as the use of "tags" to organise photographs on websites such as flickr[17] or the development of Folksonomy (Vanderwal, 2007) to categorise weblog posts into user defined taxonomies. Despite the seemingly infinite fluidity of informal classification now presented on the Internet, formalised classification have been created throughout

[17] http://www.flickr.com/

history across a range of disciplines as a platform upon which to build scientific knowledge through a shared and common understanding.

These will not be reviewed here as adequate coverage already exists elsewhere within the literature (see Weinberger, (2007); Blunt, (2001); Bowker *et al*, (1999); Lakoff, (1987:92); Borges, (1975:108)). The usefulness of classifications in the science of problem solving ideally requires that they are created objectively; however, like those cognate schema discussed earlier, many historically formalised classification have been recognised to be inherently subjective categorisations, often devised by a single or small group of people. However, with many modern classifications, methods have been developed to improve upon their scientific rigour, be this through the use of the kinds of automated clustering algorithms explored in this thesis, which objectively seek groups within large multidimensional datasets (See Everitt, 1974), through the introduction of representative scientific review bodies to oversee classification amendments (e.g. International Astronomical Union[18]), or the open dissemination of all data and methods which have been used to construct a classification (Vickers and Rees, 2007). Within this framework of best practices, classifications are used in this thesis to provide an organising concept; however, for these classifications to be effective in the context of applications to Higher Education they must assimilate appropriately defined concepts from existing education theory.

3.2 Classification and Educational Concepts

The purpose of formalised classification is to improve our shared understanding of the world by providing a simplified organising framework which helps us understand the complexity of reality. However, "knowledge about how the world works is more valuable than how it looks" (Longley *et al,* 2005:13), and as such, it is important to embed those behaviours which are classified within an encompassing

[18] The International Astronomical Union (IAU) was founded in 1919. Its mission is to promote and safeguard the science of astronomy in all its aspects through international cooperation. Its individual members are professional astronomers all over the world, at the Ph.D. level and beyond, and active in professional research and education in astronomy. (http://www.iau.org/)

and appropriate explanatory framework. Within Higher Education and educational research more broadly, human and social capital have been used to explain and explore why certain recorded behaviours may become manifest.

3.2.1 Higher Education and Capital

In the modern socio-spatial classifications that will be introduced in Section 3.4 individuals are aggregated using their home locations into a typology based on the average characteristics of the people within areas in which they live. Coleman (1988:S109) discusses how an individual's "background is analytically separable into at least three different components: financial capital, human capital, and social capital". Of these capital constructs, financial capital is perhaps the simplest to address within socio-spatial classifications as it represents the relative advantage gained by possessing increased income or wealth. Using a broad definition of income and wealth financial capital appears to be accounted for very well in commercial socio-spatial classifications where numerous consumption data are used as input variables. For example, in the literature supporting the Mosaic classification from Experian (Experian, 2007) the following variables are included:

- Credit Behaviour
- Bad Debt
- Shareholdings
- Directorships
- Property Value

However, there are complexities in linking behaviours resulting from differential financial capital if distinction is made between income and wealth. Income broadly relates to the funds gained through receiving a salary, however after those deductions relating to living expenses these become disposable income. Wealth is a function of the total value of the assets that an individual owns. An ex-housing association tenant may have bought their home from the local authority in the past, and through inflation they may now own an asset which gives them a high level of wealth: however, their income may have remained moderate. Furthermore, Friedman (1957) hypothesises that individuals may possess a concept he termed as "permanent income ". This related to how income and consumption behaviour can be balanced between both permanent and transitory income. Transitory income is the fluctuating income people may receive weekly, monthly or annually. However, permanent income is based on those long term earning projections that an individual may expect to receive. Friedman (1957) argues that consumption behaviour only changes if an individual believes their permanent income will change over the long period, i.e. independent of transitory income. Thus individuals can have varying degrees of wealth, income, and disposable income, and the combination or balance of these concepts will influence how they may behave in a given situation.

Coleman (1988:S100) defines human capital as follows:

> ***"human capital is created by changes in persons that bring about skills and capabilities that make them able to act in new ways".***

This concept is more difficult to relate to socio-spatial classification as it does not directly link to tangible assets, e.g. household income. However, human capital can be considered as an enabling framework through which greater financial capital can be accumulated. In the context of attending Higher Education, differences in human capital accumulation might be accommodated within a socio-spatial classification using indicators of variable educational attainment in post-16 qualification. Reay *et* al (2005: 21) highlight through an extensive study of access to Higher Education that "an individual's ability to deploy knowledge, skills and competencies is powerfully classed". Therefore, it is likely that between group differences in both physical and human capital might be used to identify groups that are able to leverage different advantages, to different extents, and in different ways.

Social capital is a further type of capital to which Coleman ascribes a more complex definition:

> ***Social capital is defined by its function. It is not a single entity but a variety of different entities, with two elements in common: they all consist of some aspect of social structures, and they facilitate certain actions of actors-whether persons or corporate actors-within the structure. Like other forms of capital, social capital is productive, making possible the achievement of certain ends that in its absence would not be possible.***

Coleman (1988:S109)

Thus social capital "refers to features of social organization such as networks, norms, and social trust that facilitate coordination and cooperation for mutual benefit" (Putnam, 1995: 66). Reay *et* al (2005:21) describe social capital as "generated through social processes between family and wider society and [...] made up of social networks". Thus, between socio-spatial classification groups which pertain to different levels of social capital, different behaviours could result through the exploitation of these social networks which enable members to leverage greater social returns. This is exemplified by Edwards *et al* (2003:20) who contend that "the literature relating school choice to social networks suggests that the switch from administrative allocation to a `parental choice' system extends the role of social networks, without necessarily helping to build social capital further. The reproduction of class and community differences is enhanced by the ability of parents to use their power in the education market to shape their children's future milieu".

Cultural capital is a related concept attributable to Bourdieu:

> ***"The notion of cultural capital initially presented itself to me, in the course of research, as a theoretical hypothesis which made it possible to explain the unequal scholastic achievement of children originating in different social classes by relating academic success, i.e., the specific profits which children from the different classes and class fractions can obtain in the academic market, to the distribution of cultural capital between classes and class fractions".***

(Bourdieu, 1986: 243 quoted in Reay *et al,* 2005)

Thus, in socio-spatial classification, groups with high cultural capital may express certain advantages over those groups with lower cultural capital under specific conditions. For example, entry to university to study a degree in Music may require knowledge of classical pieces with which a student from a socio-spatial group with high cultural capital may have greater familiarity. Swartz (1997:76) describes an ethnographic study conducted by Bourdieu in the French school system, stating that "French school teachers reward good language style, especially in essay and oral examinations, a practice that tended to favour those students with considerable cultural capital who in general are from privileged family origins". In a UK context Sullivan (2001) demonstrated that cultural capital transmitted within the family home has significant effect of performance in GCSE examination results. However, Sullivan (2001:893) also suggests that once cultural capital had been controlled for in her study, social class largely explained variations in attainment, and as such "cultural reproduction can provide only a partial explanation of social class differences in educational attainment".

Bourdieu and Wacquant (1992) argue that the various forms of capital can be built together, e.g. social and cultural capital could be accrued by purchasing a private education. The process of stratification and exchange can be summarised by a further quote from Bourdieu:

> ***"Those with lots of red tokens and few yellow tokens, that is lots of economic capital and little cultural capital will not play in the same way as those who have many yellow tokens and few red ones... the more yellow tokens (cultural capital) they have, the more they will stake on the yellow squares (the educational system)".***

(Bourdieu, 1993:34 quoted in Reay et al, 2005:22)

The existence and transformation of the various constructs of capital aim to satisfy the achievement of certain criteria which are "believed to be instrumental in facilitating or blocking the achievement of goals" (Rosenberg, 1956 in Krech,1962:181). Thus, the aggregate behaviours of people as recorded in empirical analysis of large datasets are related to the differentiation in common attitudes that exist within socio-spatial classification groups. It would be presumptuous to imply that all people within society share common aspirations to improve their relative position in the social hierarchy as generally perceived and measured (e.g. by income, size of house or location), yet there is a body of literature which examines social mobility within this context. For educational markets this is a highly relevant topic as it relates directly to our ability to identify those groups in society that may miss out on certain life chances through a restricted ability to compete for the advantages sustained from a quality education.

Social mobility is related to the extent that different groups within society can move upwards within or between the various "networks" and "structures" referred to by Putnam (1995), Coleman (1988) and Reay (2005) in their definitions of social capital. Sometimes referred to as intergenerational mobility, it is defined by the "extent to which a person's circumstances during childhood are reflected in their success in later life" (Blanden *et al,* 2005:2). An individual's position within such a "hierarchy" of groups is highlighted in some of the earliest definitions of social class, as exemplified in a well known quote by Max Weber:

> ***"A 'social class' makes up the totality of those class situations within which individual and generational mobility is easy and typical".***

(Weber, 1920: 302).

This quote implies that people exist within a "class" context, and mobility into a different "class" is difficult and atypical. The mapping of these concepts into the context of school and university education has been illustrated by a recent Sutton Trust report[19]. This 2007 study entailed surveying 500 people at the top of their fields in Law, Medicine, Journalism, Politics and Business. Elliot-Major (2007) found that 53% of these leading figures across the five domains had been educated at independent schools, which account for just 7% of school-age children educated in this way nationally. Furthermore, of those who had been to university in the UK, 47% had studied at Oxford or Cambridge (Oxbridge) Universities. Elliot-Major (2007) suggested that "[w]e are still to a large extent a society divided by wealth, with future elites groomed at particular schools and universities, while the educational

[19] http://www.suttontrust.com/

opportunities available to those from non-privileged backgrounds make it much more difficult for them to reach the top." The extent to which industry structures are overrepresented by those who have received the most privileged education provides evidence not only of differentiation by "wealth" but more broadly the possible influence of social, human or cultural capital. Thus, it is suggested that possession of increased financial capital broadens school choice, whether through the ability to pay for a private education or freedom to move into areas in close proximity to the best schools. Lampl (2007) suggests that "[t]he first priority should be to improve our underperforming state schools but we also need to recognise that we have a socially selective school system. [...] The top 20% of our secondary schools - independents, grammars and leading comprehensives - are effectively closed to those from non-privileged backgrounds", which presents issues of social mobility for those living in less affluent neighbourhoods. A further report commissioned by the Sutton Trust on social mobility concluded "[i]ntergenerational mobility fell markedly over time in Britain, with there being less mobility for a cohort of people born in 1970 compared to a cohort born in 1958 [...] part of the reason for the decline in mobility has been the increasing relationship between family income and educational attainment between these cohorts. This was because additional opportunities to stay in education at both age 16 and age 18 disproportionately benefited those from better-off backgrounds" (Blanden *et al,* 2005:2). An alternative view to this study has been presented by Smith (2007) who suggests that some of the empirical findings may be misleading, and that the study actually "suggest quite a high degree of social mobility". For the two cohorts of people that were tracked, in the study Smith (2007) suggested that the report actually illustrates that in the UK 63% of those born into the lowest quartile of households as measured by income had progressed up the scale, and that in the 1958 cohort this was a similar figure of 69%, marginally better than in 1970. Thus, he suggests that the results actually show a much lower reduction in social mobility than presented by the conclusions of Blanden *et al,* (2005). Smith (2007) also makes the point that in the 1950s "class" was more closely associated with the grades in an economy led by manufacturing, and that social gradation was then more about manual versus administrative functions. Thus, the Weber (1920:302) notion of class introduced earlier where "individual and generational mobility is easy and typical"; this refers directly to those more contacted employment opportunities available at the time. Thus, it would be highly unlikely for a son of a factory worker to become a lawyer for example. At this time assessment of progression between groups would be reasonably obvious, however in the post industrial society with a fragmented jobs market these differentiations become substantially more difficult to interpret (Longley and Webber, (2003)). Thus within this context of social measurement the following section considers how those concepts relevant to policy issues and outcomes in Higher Education may be measured.

3.3 Social Measurement, Classification and Indicators

Social measurement relates to the construction of both classification and indicators which can be used to assess changes in society. Whereas classification implies a multivariate categorisation of objects into discrete boundaries or hierarchy, indicators differ in that they attempt to measure rates in specific conditions either directly or indirectly. Indicators can be either univariate or multivariate in their composition. For example, changes in population characteristics could be measured directly by an individual level national census or indirectly through inferences derived from various data sources such as GP registers, household waste volume or school registrations. However, classification and indicators are not mutually exclusive and as discussed in the previous section indicators are sometimes used as input into socio-spatial classification such as those indirect measures of "wealth" that Experian use as input into its Mosaic classification. Additionally, classifications are sometimes used to create local indicators of predicted behaviours (Harris *et al*, 2005). The more complex a multidimensional classification, in terms of the mixture and types of indicators that it contains, the diversity of its sources, the different scales of measurement and the varying spatial resolution that it entails, can all raise issues that are problematic in public sector applications and increase the need for methodological openness.

3.3.1 Multivariate Indicators in the Public Sector

Considerable research into the classification of populations for public sector applications has been carried out in the past. There has been work in health on the creation of numerous multivariate deprivation indices, with a range of applications in targeting provision and weighting performance. Indicator measures include the Jarman Index (Jarman, 1983), Townsend Scores (Towsend and Beatie, 1988) and the Carstairs Index (Carstairs and Morris, 1991). In each of these classification there is a priori reasoning to support the linkage of a concept (specifically, in these instances, the relationship between GP work load and deprivation) to the different sources of data used in construction of the indicator. Each of these classifications is a composite of different Census variables, scaled and grouped in order to measure dimensions of deprivation. The Jarman index, for example, is created from the following variables:

- Unemployment
- Overcrowding
- Lone Pensioners
- Single Parents
- Born in New Commonwealth
- Low Social Class
- One Year Migrants

The exact calculation using the 1991 Census variables by their standard codes can be found on the Census Dissemination Unit (CDU) website[20]. The original policy objective for the index was to assign extra funding to reflect the workload of general practitioners in deprived areas; however the ability of the index to achieve these aims has been questioned (see Carr-Hill and Sheldon, 1991). Further criticisms of the Index include the spatial variability in performance (Talbot, 1991) or the

[20] www.census.ac.uk/CDU

comparisons of the classification to the performance attained using direct socioeconomic information (Marsh *et al,* 2000). Townsend scores were developed for the Northern Regional Health Authority as a further measure of material deprivation, and use the following Census variables:

- Employment
- Overcrowding
- Non Car Ownership
- Non Home Ownership

Although Marsh *et al* (2000:630) found "the association between Townsend score and health status was strong enough to be of practical importance", the variables included in creating the summary score should only reflect aspects of deprivation which are uniform across all areas. Car ownership in a dense urban area such as London may not indicate deprivation because other modes of transport are more prevalently used. The Carstairs index was developed subsequently from the Townsend scores by the analysis of Scottish health data. This new classification was sought to reflect certain elements of deprivation which were believed to be unique to Scotland. The following variables are included in the Index:

- Unemployment
- Overcrowding
- Non Car Ownership
- Low Social Class

Since the year 2000 two re-workings of a related, more generic, deprivation indicator have been created. The most current of these is the 2004 Index of Multiple Deprivation (IMD)[21]. The IMD is disseminated at Super Output Area and is designed for a range of public sector applications. The classification is created from 7 different domains including:

- Health deprivations and disability
- Employment
- Income
- Education, skills and training
- Living environment
- Barriers to housing and services
- Crime

In addition to scores and rankings for these 7 domains an overall score is created as a measure of general deprivation. Although freely available to use, there is little evidence that either the aggregate classification, or the education domain have been used in peer reviewed education literature to the extent that is has been adopted in health. HEFCE has made some effort to introduce spatial classifications into its funding models with the introduction of the Participation of Local Areas (POLAR) classification in 2003[22]. This classification was derived from a demographic study of young (18 – 19 year old) participants in Higher Education between 1994 and 2000. Rates were derived at the Ward level and these scores disseminated as a series of quintile bands.

[21] The IMD is supplied by a department of the government called Communities and Local Communities (Formally the Office of the Deputy Prime Minister). The classification is available from: http://www.communities.gov.uk/index.asp?id=1128440
[22] The full methodology for constructing POLAR can be found at: http://www.hefce.ac.uk/pubs/hefce/2005/05_03/

3.3.2 Some Examples of Measurement in Education

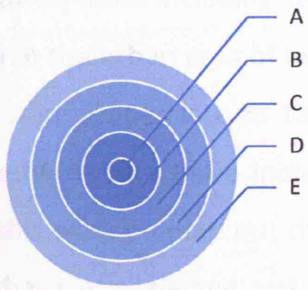
In educational applications there are various methods which are or have been used past and present to transform data into information (Longley *et al*, 2005) including classification, direct and indirect indicators. Investigations into Higher Education participation have most prevalently used individual level occupational classification. Archer *et al* (2003) review these classifications to include the National Statistics Socio Economic Classification[23] (see also Rose and Pevalin, 2003), the Socio-economic Group and the Registrar General's Classification. Like the early conceptions of class (Weber, 1920) these formal classification group individuals according to their occupational category. For applications to Higher Education this should cause concern, as when assigning these variables before the age of 21, UCAS assigns occupational categories based on applicants' parental occupation, whereas after the age of 21 these are assigned on the basis of applicant occupation. Thus, within this classification parental occupation is used as an indirect indicator of the student's background, whereas a more appropriate measure could perhaps be derived using a neighbourhood level classification that directly reflects the milieu in which the individual grew up. Since 1994 the UK Government has published school performance tables for secondary school GCSE and A-Level results, which additionally were supplemented with primary school attainment data in 1996. These school performance data directly measure attainment in exams taken by those students within schools and allow comparison to be made against equivalent performance within the local authority and also the national average. National coverage school data do not include parental occupation and as such in educational analysis socio-economic differentiation is often implied through analysis of free school meal rates within schools as a substitute for a level of deprivation (See Shuttleworth, 1995). In Higher Education performance indicators are constructed for widening participation, retention rates, research output and employability of graduates. These are created from a range of both direct and indirect indicators taken from those attributes collected on students studying within Higher Education in the previous academic period.

[23] http://www.statistics.gov.uk/methods_quality/ns_sec

3.3.3 Static Classification and Dynamic Processes

Social measurement has been discussed thus far in terms of static states of behaviour. However, these measured events do not occur in temporal isolation, and as such social measurement can be considered by the extent that classification or indicators capture these dynamic processes.

This is a relevant concept to the urban land use models developed in early 20th century Chicago. Robert Park and colleagues applied ecological concepts to explain observed segregation and patterns of social interaction within Chicago. Knox (1987) discusses how these 'natural' processes included dominance, segregation, impersonal competition and succession. Park proposed that the fundamental principle of competition between individuals occurred through the operation of markets and resulted in patterns of land use and rent, and as such, the segregation of people into distinctive areas. Burgess (1925) used these ideas to model the processes of neighbourhood differentiation, developing his famous concentric ring model (see Figure 3.3). In this model the city grows outwards and differentiation between changing zones is established by the ecological processes of invasion and succession, or in other words expansion and land use conversion. The contribution of these studies to the history of urban measurement lies in the paradigm shift that it stimulated away from focused research on specific areas to the establishment of more general statistical relationships between urban systems and their structure (Harris *et al*, 2005).



A=Central Business Districts, B= Zone of Transition, C= Zone of Factories and working men's homes, D= Residential homes, E=Outer commuter zone.

Figure 3.3: The Burgess Land Use model

Further models that developed out of this same school of thought included Hoyt (1939) who proposed a sector model that was based on land use being divided by a series of wedges that radiate from the CBD to account for the effects of communication lines such as roads (See Figure 3.4).

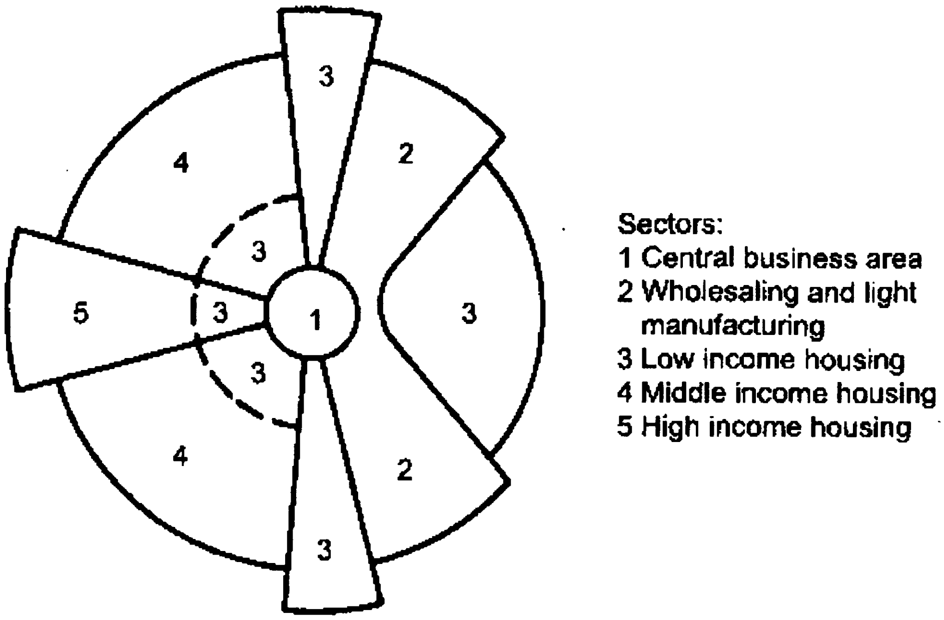


Figure 3.4: The Hoyt Sector Model (Source: Hoyt,1939 in Harvey, 2000)

However, the Human Ecology school subsequently became discredited as the models were applied to other locations outside the case study cities, and as Knox (1987:61) discusses they were later 'abandoned in favour of a much modified ecological approach based on the idea of identifying key variables and examining their relationships within an ecosystem or ecological complex'. These social area analyses were developed by Shevky and Williams (1949) and involved creating a series of scores for a number of variables including social rank, segregation and the degree of urbanisation, all within the urban area of in Los Angeles (LA). Using these three indices, LA census tracts were divided into an 18 Group classification based on how similar the areas were between the three indices. Bassett and Short (1980) discuss how initial reviews of this work pointed out that there was little theoretical justification for the variables that were selected and the indices that were created using them. What became known as social area analysis was later re-presented by Shevky and Bell (1955) as a deductive model: however Bassett and Short (1980: 17) note that "social area analysis as theory is little more than a retrospective rationalisation for previous empirical work". Thus these early models, although a progressive step forward in urban theory were essentially created through a process of inductive theory building, which is "theorising from a mass of observations" (Wilson, 1971:32) and therefore illustrate a weak linkage between concepts and measurement. Thus, it is important for modern measurement techniques to be grounded in theory and can be validated through repeated observation.

3.4 Geodemographics, Socio-Spatial Differentiation and Change

Geodemographics represents a multidimensional technique to measure socio-spatial differentiation and change extending as a progression from those urban models in the "ecological tradition" (Bassett and Short, 1980:9). These techniques are essentially deductive models of urban processes which build from those models discussed in the previous section and additionally relate to work Booth (1889) undertook to study poverty in London. Booth classified streets in London by the prevailing socio-economic characteristics of the areas, assigning them into one of 7 groups which were colour coded and link to a series of descriptions which represented a range of social conditions, specifically poverty (See Figure 3.5 and Table 3.1).

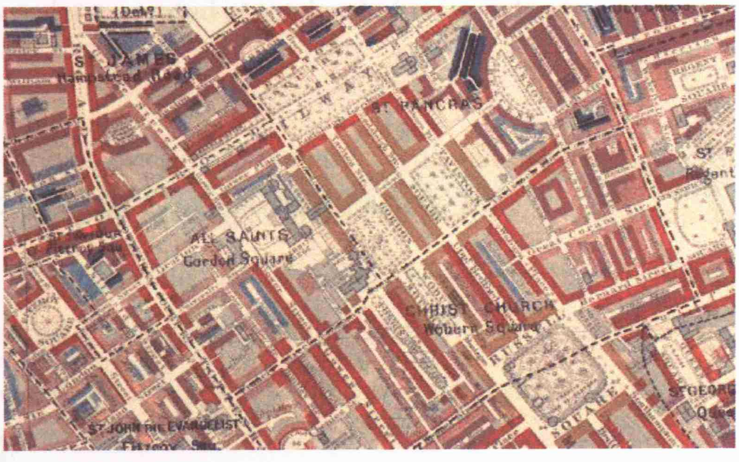


Figure 3.5: A Booth Map of the Area Surrounding UCL (Source: Booth Maps Online[24])

[24] A searchable version of the Booth map is available online from: http://booth.lse.ac.uk/

Table 3.1: The Booth Map Key

Colour Code	Description
	BLACK: Lowest class. Vicious, semi-criminal.
	DARK BLUE: Very poor, casual. Chronic want.
	LIGHT BLUE: Poor. 18s. to 21s. a week for a moderate family
	PURPLE: Mixed. Some comfortable others poor
	PINK: Fairly comfortable. Good ordinary earnings.
	RED: Middle class. Well-to-do.
	YELLOW: Upper-middle and Upper classes. Wealthy.

Harris *et al* (2005) discuss how the clustering of small areas into groups using numerous census variables originated in the late 1970s as a deprivation classification of wards, parishes and enumeration districts in Liverpool. The Office of Population Censuses and Surveys (OPCS) commissioned this research from the Centre for Environmental Studies in London, based on previous small area deprivation studies in Liverpool (Sleight, 1993). Richard Webber is credited with much of this early research (see Webber, 1977; Webber, 1978; Webber and Craig, 1978). This research was eventually bought by the company CACI who introduced these methods into the private sector under the brand ACORN (A Classification of Residential Neighbourhoods). These data on socio-economic characteristics were thus used to imply consumption characteristics relevant to the private sector. The classification was later created at the level of unit postcode.

These classifications potentially commit ecological fallacy (Robinson, 1950) by design, since they seek to predict individual consumer behaviour using indicators pertaining to areal aggregations. Tranmer and Steel (1998) contend that these ecological effects occur despite individuals who live in close proximity, however the strength of these problems depends on the exact area of aggregation being studied (Martin, 1991).

Furthermore, geodemographic analysis may suffer problems with the Modifiable Areal Unit Problem (MAUP) (Openshaw, 1984) which occurs when the mean attribute values of geographical areas change depending on the number of areas into which a population is divided (Tranmer and Steel, 1998) or when zonal boundaries are moved or modified. Separately or together these scale and zoning effects form the MAUP. This is of potential concern to the construction of geodemographic classifications, since some are built as aggregations from a range of geographical units and none is founded exclusively upon observations pertaining to unique human individuals (Fotheringham and Wong, 1991). This is most recently demonstrated in a geodemographic context by Experian releasing the Mosaic classification for academic use into the National Data Archive. The data supplied were aggregated up from unit postcode to district level. However, Richard Webber (founder of modern geodemographics) notes in De Smith *et al* (2007: 96, cited as personal communication) as response to criticisms of Geodemographics in relation to MAUP that:

> ***"I have never to come across any real world example of a conclusion being invalidly reached as a result of this hypothetical possibility"***

Webber in De Smith *et al* (2007:96)

There has been little attempt within the literature to identify the relationship between those concepts of capital formation introduced in Section 3.2.1 and geodemographic analysis, perhaps with the exception of Webber (2007). Unlike "class" which is a concept traditionally used and understood in the broad sociological literature, geodemographic typologies have only recently begun to seek to provide, or perhaps post rationalise, a conceptual framework to understand reality (see Parker *et al* 2007). The experience of Shevky and Bell (1955) suggests that we have been here before. Drawing parallels between concepts of social stratification derived from sociological literature and labels used in the Mosaic typology, Webber (2007:185) suggests that these "incorporate language which corresponds closely to that used in the discourse on both globalisation and gentrification, and studentification", although he cites only a single source (Atkinson and Bridge, 2005) in making these assertions. The shorthand labels and pen portraits (Birkin, 1995) assigned to the elements of different classes within any typology are undoubtedly useful, yet vivid description does not of itself provide academic rigour, particularly when the methodology for the construction of the commercial classification is closed to external scrutiny. There are indeed worrying similarities between the criticisms that Skevky and Bell (1955) sustained in relation to their post facto rationale of social area analysis and the arguments presented by Webber (2007) in support of geodemographics. Geodemographic analyses clearly have an inductive legacy, and to retrospectively argue that current typologies are founded upon ad hoc strands of social theory is academically misleading. A different route is for classification vendors to re-create future classifications from a more theoretically informed and deductive standpoint.

Within the private sector, geodemographic analysis has conventionally been used to examine both incremental purchases (such as those measured by newspaper readership) and discrete consumption (such as the propensity to consume holidays of different types) of private goods. In principle there seems no reason not to anticipate the suitability of these techniques to investigate aspects of consumption of public sector goods and services. However, government statistics have traditionally used data derived from the census, administrative records and sample surveys to allow them to meet the majority of their needs for services targeting in local areas (Longley and Webber, 2003), and for this reason geodemographic analysis has to date been less prevalent within public sector applications. There has been a growing number of applications in health (see Aveyard *et al*, 2000; Tickle *et al*, 2000; Stafford and Marmot, 2003), crime (Massimo *et al,* 2001; Bowers and Hirschfield, 1999, Ashby and Longley, 2005) and education (Tonks, 1999; Tonks and Farr, 1995) that explore the use of GIS and geodemographic analysis to assist in policy and decision making at both the local and national levels. These studies have developed in parallel with government reforms of public services, and indeed the initiative "Big Conversation" which encouraged discussion on replacing a one size fits all welfare system with one of 'individual aspiration backed up by strong communities' (Blair, 2003). This shift in the focus of public policy towards individuals and the neighbourhoods in which they live are driving micro level data requirements, and to some extent mirror the developments seen in the private sector.

A further strand to these shifting interests stems directly from government policies that have encouraged public sector organisations to behave in ways akin to those of the private sector. These convergences can be illustrated by the replacement of government Compulsory Competitive Tendering Policy with the concept of "best value" for the repeat purchase of public sector goods and services. This policy is consistent with public-private collaboration, and as the geodemographic industry has been shown to be successful and highly profitable in the private sector (Harris, 1998), its adoption in the public sector seems to be a logical and progressive step forward. This becomes of growing importance as the geodemographics industry becomes increasingly open to competition in the UK, not least as 2001 Census data are now freely available to download, allowing end users the ability to create their own systems or 'value added' by incorporating their own sector specific data. Prior to the 2001 Census this was also possible but only by being tied into expensive licence agreements with the small number of census distributors who in turn licensed public sector data from the Office of National Statistics (ONS)/ Office of Population Censuses and Surveys (OPCS). One such development has led to the creation of the ONS Output Area Classification (OAC: Vickers and Rees, 2007) which has an entirely open and public domain geodemographic classification. This classification is seen as a positive step forward; although without the underlying data being refreshed (as in commercial classification) there is danger that this classification could become out of date.

3.4.1 Privacy and Ethics

When attempting any form of categorical classification of areas or individuals, there is the underlying assumption that reality can be accurately represented by the typologies, and debates and dialogues justifiably extend to how the groups were conceived, measured and otherwise created. For classifications used in the public sector this is of particular significance, especially if the application of techniques may affect real life chances of those who are classified correctly or otherwise. Questions concerning the integrity underpinning the use of geodemographics in the analysis of social processes and stratification often cite Harvey's (1973) concern that these techniques develop knowledge that appear to be true, but that in actuality hide the truth of reality. Furthermore, Sui (1998:662) suggests that these 'instrumental approaches generally take an atomistic ontological position in which the social position of the researcher is independent of the knowledge that he or she produces'; suggesting that data led empirical investigations may not be sufficient to represent the complexity or dynamics of real world social processes.

Geodemographic information systems have also been criticised as they threaten privacy in two key ways. First, Goss (1995) describes how mis-specification problems with a database can inadvertently discriminate even if the use of the data is legitimate in de jure terms. In the context of Higher Education these effects could be particularly acute as the consequences of mis-specifying disadvantage in terms of educational services has potentially very serious consequences, and particularly so if life chances are being apportioned. It is therefore essential that, if these indicators are adopted, analysis be conducted into whether they will ameliorate or compound these issues. Goss's (1995) second privacy concern is that data on an individual gathered for one purpose may be transferred to another context without that individual's permission. "Off the Shelf" geodemographic indicators are constructed with legally available data without infringement of this aspect of the Data Protection Act, thus negating this second concern, albeit only in a strict legal sense. However, when these indicators are appended to other data such as university application successes, extra caution must be taken to ensure that the spirit of these safeguards is adhered to. Taken together, it is clear that when using either geodemographic or socio-economic indicators it is important to ensure that the context of the investigation is understood, in order that erroneous interpretations may be avoided. For example, in an investigation into Higher Education participation using the Mosaic geodemographic classification, it may be inferred that an area classified as "Welfare Borderline" has lower Higher Education participation rates because of the restriction created by prevailing lower incomes (assuming that this was identified as a key variable for lower participation). Yet in reality there may be other important contextual factors in these areas such as the types of employment, and it is these that may affect the weighting or importance that is placed upon Higher Education and attitudes to social capital formation. Resulting human capital formation from participation in Higher Education is unlikely to be a simple function of income alone and it is more likely to accrue through a combination of socio-economic processes.

3.5 Measuring and Modelling Educational Choices and Decisions

The usability of a classification schema in measuring or modelling educational choices is directly related the homogeneity of the composite groups of applicant / institution choices and decisions. Individual participation choices fall within the broad categories of whether, where and what Higher Education courses attract the participation of an individual. These choices then intersect with the decisions that an institution makes, such as offer or rejection. In this context education can be considered a commodity, and applicants as consumers. Hensher and Johnson (1981:12) discuss that the 'selection decision between commodities which are perceived to be available' is influenced both by physical and aphysical factors. Physical effects on demand could be course grade requirements, and aphysical factors could be that a candidate may perceive a particular institution as 'too posh', or as 'ivory towered' (Singleton, 2003). Choices may be deemed to entail 'selection decisions by an individual between commodities which are perceived to be discrete and which are contained in a relative choice set' (Hensher and Johnson, 1981:11). Commodities are defined at either the intensive or extensive margins. Higher Education provision falls at the extensive margin, as an increase in cost cannot usually result in a partial reduction in consumption in the same way as with a continuously defined product such as the purchase of meat from a butcher. A consumer, in this case applicant, will choose (or not) a course of Higher Education that maximises their utility from a choice set by consuming different levels of attributes from choice alternatives, where utility is the 'index of the relative levels of satisfaction associated with the consumption of particular commodities' (Hensher and Johnson, 1981:11). In Higher Education these attributes could include qualification aim, price of the course, or reputation of the institution. An applicant will make joint course and institution choices according to a set of value judgements that maximise perceived personal benefit over cost. Factors that may influence these choices might include:

- Residential location
- Subjects offered and course structure
- Facilities
- Grades
- Reputation
- Parents and family
- Friends
- Schools
- Price
- Bursaries

Therefore, using the maximum utility principle an individual choice set may be limited or enhanced by one or combination of these factors. For example, a single mother living in Bolton and wishing to extend her education, may have a choice set constrained to a single institution. However, a final year A-Level student, living in

Buckingham and attending a top performing school may have a much less constrained set of choices.

Institutions will also make value judgements as to which students they wish to admit in order to meet their recruitment or widening participation targets. These might include the following variables:

- Personality
- Interview performance
- Grades and subjects
- Entrance tests
- School type
- School Achievement

The interaction between student and institution choices creates a distribution of acceptance to courses of Higher Education. The effectiveness of a geodemographic classification for use in Higher Education might therefore be a function of its tacit ability to measure the variables which influence these behaviours. As part of a major 39 month research project into educational choices which consisted of around 137 in depth interviews Ball *et al* (1996: 104) discuss how socio-spatial differentiation in education arises and is reinforced as a function of the choice process available to different segments of society:

> ***"…choice is very directly and powerfully related to social-class differences …choice emerges as a major new factor in maintaining and indeed reinforcing social-class divisions and inequalities".***

Ball *et al* (1996: 104)

Ball *et al's* (1996) use of "maintaining" and "reinforcing" in this quote links with the role of education in restricting or enhancing social mobility as shown in literature reviewed in Section 3.2.1. The apportionment of different constructs of capital between people in society creates variability in the choice sets available to them. Gewirtz *et al* (1993) introduced an organising framework for these choices of *inclination* and *capacity*. Tooley (1997:218) summarises these two factors as:

> ***"Capacity itself has material and cultural dimensions. Material capacity includes the resources to pay for transport to and from school (including private cars and taxis), improved housing, extra-school coaching, private school fees and child-care opportunities. Cultural capital capacity includes having knowledge about and familiarity with the education system, self-confidence, and 'stamina-to research, visit schools, make multiple applications and appeal'***

Tooley (1997:218)

> ***"Inclination involves the extent to which they were inclined to be engaged with the choice system; higher inclination implies possessing certain 'beliefs' about schools, for example, that they differ in terms of: atmosphere, the 'standard of education' they offer, their exam results, the life chances they facilitate, the values they impart, the extent of extracurricular activities, the kind of children that go to them, levels of resourcing, levels of parental support, and the commitment of teachers".***

Tooley (1997:218)

From the literature reviewed in Section 3.2.1 relating to capital, Tooley's (1997) economic capacity directly relates to concepts of economic capital, and cultural capacity to cultural capital. However, inclination is a more complex attitudinal construct and relates to the engagement and motivation of different groups which make up society to make educational choices. Furthermore, these conceptions of educational choices relate to earlier discussion on utility maximisation (Hensher and Johnson, 1981). Those groups of people as defined by a geodemographic typology are going to have variable capacity and inclination, thus placing them within the Gewirtz *et al* (1993) choice matrix as presented by Tooley (1997:219) at different locations (See Figure 3.6).

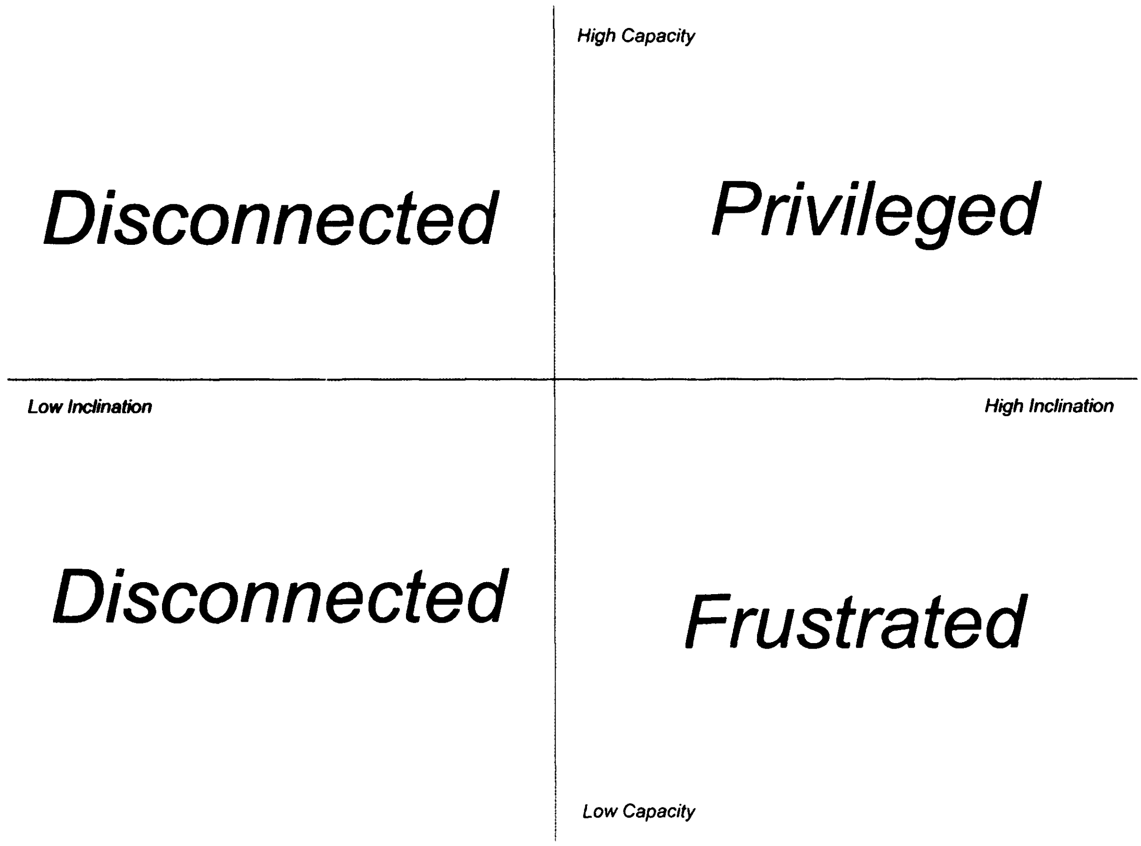


Figure 3.6: Matrix of Choice Characteristics

3.6 Conclusion

The emergent contention of this wide-ranging review is that for a classification to effectively differentiate in educational applications, it must seek to measure the behavioural differences arising from choice restrictions exhibited through variable inclination or capacity. In order to ensure that the concepts discussed throughout this chapter are effectively mapped into homogenous classification categories, there is perhaps a requirement for bespoke educational classifications or indicators which integrate these behaviours directly through use of the underutilized data that were discussed in Chapter 2. From this discussion, the interim conclusion is that it appears that general purpose classification systems (such as those marketed by commercial providers) can claim no particular status in accounting for the consumption of the various services provided by the *public* sector - with perhaps the single partial exception of Health ACORN[25]. This does not directly include health sector data, but includes other data relating to health outcomes, such as diet

[25] www.caci.co.uk/acorn/healthacorn.asp

information. However, the implied assumption that that the nature of individual use of public services such as education should directly correspond with the ways in which consumers uses private goods and services is problematic if not fatally flawed. Finally, in order that social measurement may be incorporated into Higher Education decision making, there appears to be a requirement that appropriately constructed classifications be available through some centralised method of dissemination which enables the integration of the data outlined in Chapter 2.

PART 2: GEODEMOGRAPHIC PROFILING AND DECISION SUPPORT FOR HIGHER EDUCATION

4

THE SOCIO-SPATIAL CONTEXT TO HIGHER EDUCATION ACCESS

4.1 How can we Understand Access?

This chapter will introduce a series of univariate indicators which enable better understanding of what we mean by access to Higher Education; how distributions may manifest and be appropriately measured. This analysis aims to unravel some of the social and spatial complexities which surround access issues in Higher Education by using geodemographic classification as a multidimensional organising framework through which aggregate behavioural choices can be measured. A series of analyses are presented which relate to a number of important indicators which are highlighted by the literature to influence differential social and spatial neighbourhood access rates including distance (Harris *et al*, 2007), prior attainment (Leathwood and Hutchings, 2003), course choice (Reay *et al*, 2005) and age (Archer, 2003).

Consideration of these issues begins to develop a notion of which indicators are to be most usefully included in those tools required for decision support which will be explored in later chapters. Linking access indicators to a common framework of geodemographic classification provides stakeholders with a simple method through which those complex behaviours can be measured. Different stakeholders however will be interested in different indicators. For example, selecting institutions may be interested in those students from neighbourhoods who supply the best prior qualified candidates whereas recruiting institutions may have more concern for the distances that students from particular neighbourhood groups typically travel to accept an offer at an institution. In the analyses presented in this chapter the Mosaic classification is used as at the time this was the only classification that UCAS were licensed to use.

4.2 Problems of Propensity and Data Representation

This chapter explores the spatial and social complexities that are present in geographies of access to Higher Education as "it is generally helpful to look at a data set before any models are fitted or hypotheses formally tested" (Fotheringham *et al,* 2000:65). However, before these empirical observations it is appropriate to examine a range of methods to explore these spatial data. A common technique used in both commercial and academic geodemographic analysis is index scores.

An index score can be used to show the overrepresentation of a target group by a discrete classification when compared to its proportions in a base population – as in Equation(4.1) where index scores *I* are calculated by comparing the proportion of a variable *v* within a target population *t* relative to a base population *b*.

$$I_v = \frac{\dfrac{t_v}{\sum_{n=1}^{n} t_v}}{\dfrac{b_v}{\sum_{n=1}^{n} b_v}} \times 100$$

(4.1)

To illustrate the benefit of using index scores a number of maps will be presented using a 2001 Census variable which measures the frequency of people within Output Areas aged between 16-74 who have a qualification which are 'Level 4 or 5', e.g. a degree or masters degree. The choropleth maps were created with 10 quintiles for counts (Figure 4.1), percentages (Figure 4.2) and index scores (Figure

4.3). The percentage map was created using the denominator of the total 16-74 population and the index scores are calculated using a base derived for the whole of Buckinghamshire. All three maps show three very different spatial patterns using different representations of essentially the same data. The map showing purely frequency is effectively a representation of the total population, e.g. you would generally expect a higher frequency of a given target variable where the overall population is higher. Therefore, counts are obviously not suitable for discovering whether and if so how overrepresented characteristics are relative to a population within an area. Percentage scores partially explain overrepresentation of one variable in relation to another, standardising the local variable against a local population. However it is index scores which compare these local percentages against another population, be this regional, national or specific aggregations to demonstrate how over or under represented a variable is compared to these base scores.

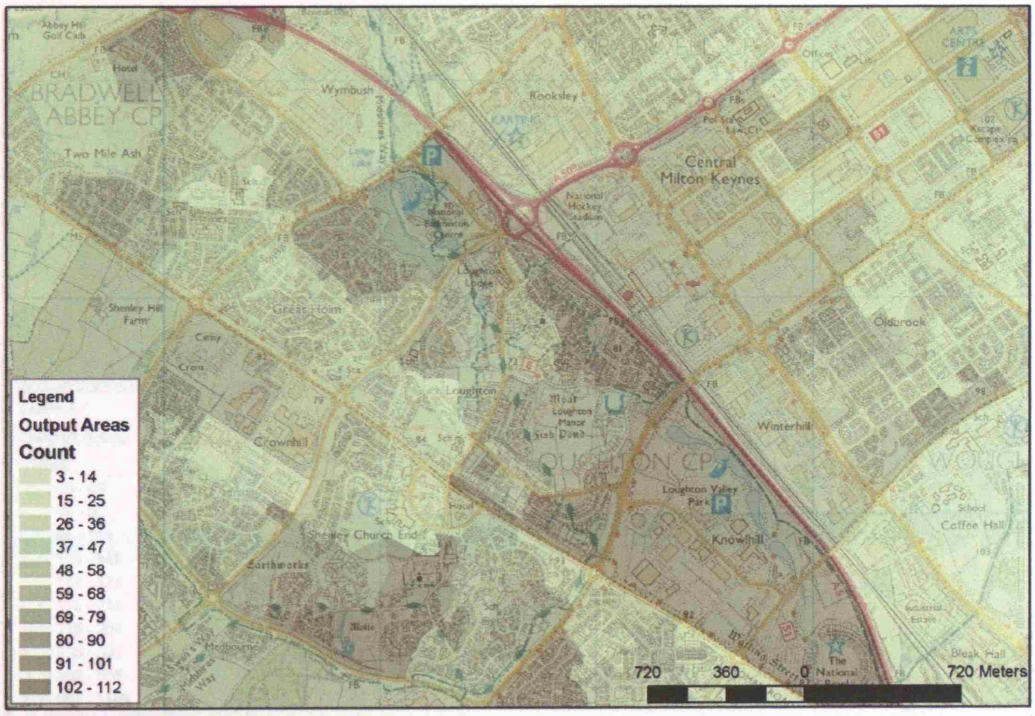


Figure 4.1: Frequency of Level 4/5 Qualifications in Milton Keynes (Source: 2001 Census)

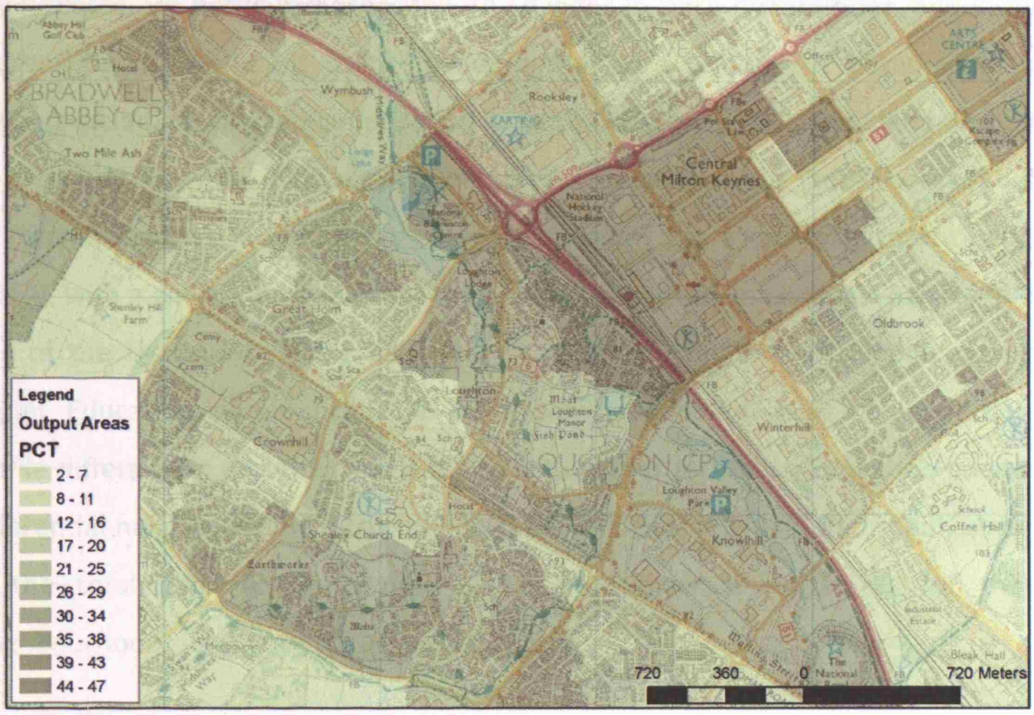


Figure 4.2: Percentage Level 4/5 Qualifications in Milton Keynes (Source: 2001 Census)

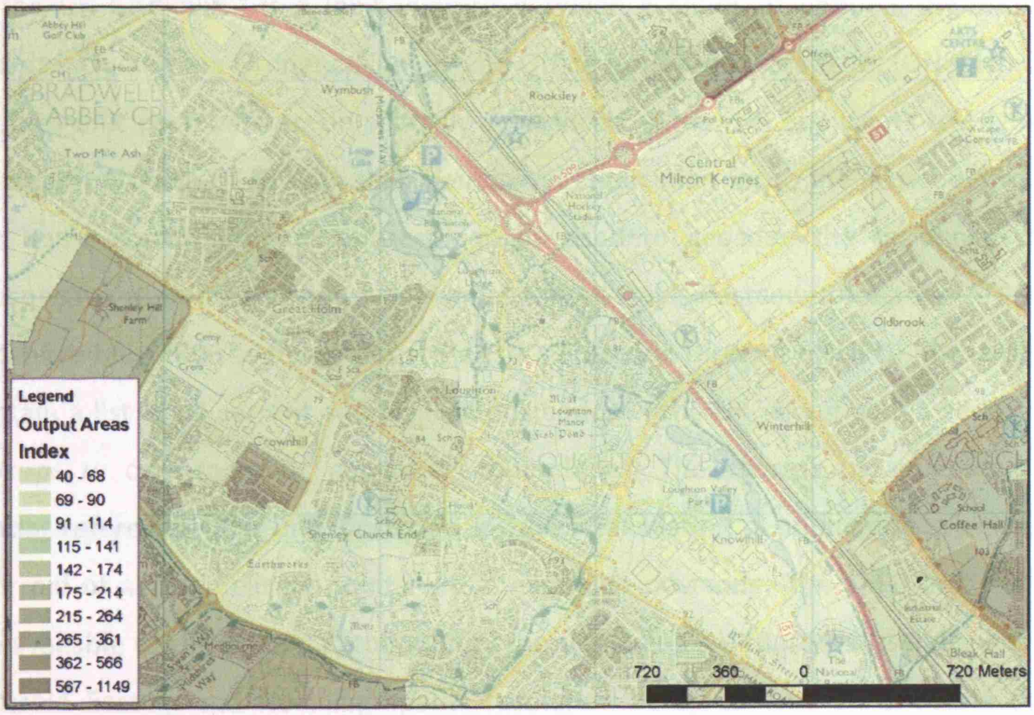


Figure 4.3: Index score Level 4/5 Qualifications in Milton Keynes (Source: 2001 Census)

Index scores are particularly sensitive to changes in base distributions, and in any Higher Education analysis using geodemographics a number of bases could be derived including all Higher Education acceptances or Census population counts which would be useful in different contexts and affects how the index scores should be interpreted. When comparing a target variable, e.g. a university participation profile, with a census derived population base this broadly measures a participation rate of the institution. If the base were calculated from the total population of Higher Education acceptances the index scores represent how an institution's profile differs from all other institutions. With access to Higher Education creating differential neighbourhood participation rates, this profile would be broadly similar to the total participation profile, albeit with lower index scores for those neighbourhood types least likely to participate.

4.3 Higher Education Choices and Distance

In the UK Gridlink®[26] is a joint venture between a series of national organising bodies including the Ordnance Survey and Office for National Statistics (ONS) which enables unit postcodes to be spatially referenced. When UK applicants to Higher Education provide UCAS with their completed application form, the postcode from the home correspondence address can be used to geocode the residences of prospective students. Previously known as the All Fields Postcode Directory (AFPD), and since May 2006 as the National Statistics Postcode Directory (NSPD), these files contain a list of all current and expired postcodes in the UK, along with a series of lookups to corresponding geographic boundaries. The process of appending a spatial reference to a unit postcode involves calculating a population weighted centroid of all the delivery points within each unit postcode. The AFPD and NSPD file available to the academic community through Edina's UKBorders[27] service has a centroid Easting and Northing spatial resolution at up to 1m. UCAS define the location of a Higher Education institution based on the unit postcode associated with the main institutional campus, and as such this can introduce some error into those institutions who have multiple sites. Using the AFPD file, 2004 UCAS degree

[26] http://www.statistics.gov.uk/geography/gridlink.asp
[27] http://edina.ac.uk/ukborders/

applicants who accepted a place were geocoded, using their domicile postcode. Location of accepting institution was defined in the same way and the Euclidean distance between these two locations were calculated for each applicant. These straight line distances between student domiciles and institutional locations *i* is calculated using Equation (4.2).

$$d = \sqrt{(x_s - x_i)^2 + (y_s - y_i)^2}$$

(4.2)

Distance scores can be aggregated and partitioned using a series of other variables available in the UCAS data. This section will examine some of these spatial complexities, demonstrating how courses, institutions and socioeconomic factors all influence the distance that an applicant is willing or able to travel to accept a place at an institution. Figure 4.4 shows the median distance travelled by each placed student according to 2001 Census Super Output Areas. The darker colours indicated longer distance travel and the red dots are the location of UCAS institutions. Where clusters of institutions occur there is a tendency for shorter distance travel, most significantly represented in London (see map inset). From rural areas, as one might expect that students generally travel further distances.

4.3.1 Neighbourhood Classifications and Distance Travelled

Although of interest to national Higher Education policy, the aggregate spatial distribution shown in Figure 4.4 does not highlight heterogeneity in distance travelled to different Higher Education institutions, to participate in different courses, or, from different neighbourhood types. Calculation of distance travelled according to Mosaic Types for all 2004 UK degree acceptances (See Figure 4.5) makes clear the high variability between Types and Groups. These patterns are partially explicable by the spatial distribution of the neighbourhood Types themselves. Accepted applicants from the Group "Rural Isolation" have a higher propensity to travel greater distances, as the areas where these neighbourhood types occur are in areas distant from the major urban conurbations where most Higher Education institutions are located. Another Mosaic Group with a high propensity to travel is "Symbols of Success" which contains the most affluent neighbourhoods in the UK and that are often found in large and prospering cities. These applicants will likely have fewer financial restrictions on their abilities to travel larger distances to accept places at an institution, and with a history of Higher Education more prevalent within the families living in these areas, a culture of moving away from home to university is more probable.

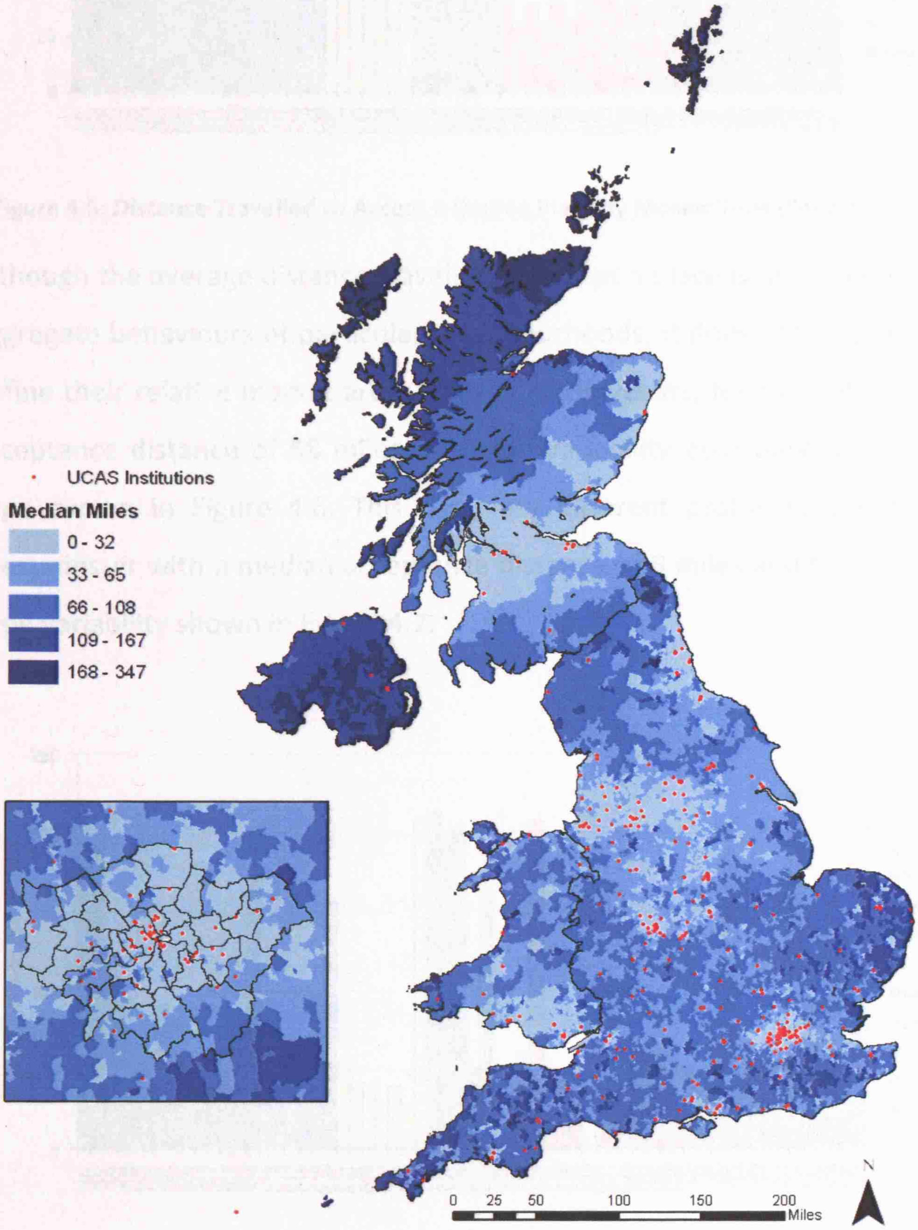


Figure 4.4: Median Distances Travelled by Students Resident in each UK Super Output Area (Source: 2004 UCAS Data)

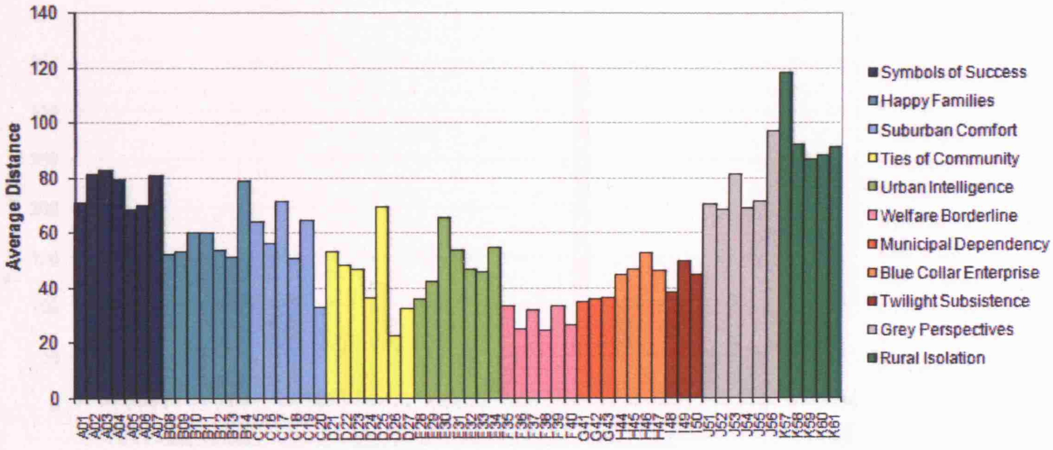


Figure 4.5: Distance Travelled to Accept a Degree Place by Mosaic Type (Source: 2004 UCAS Data)

Although the average distance travelled to accept a place is useful in describing the aggregate behaviours of particular neighbourhoods, it does not help institutions to define their relative market areas. Lancaster University, for example, has a median acceptance distance of 58 miles, with the variability according to neighbourhood Type shown in Figure 4.6. This is a very different profile to the University of Westminster with a median acceptance distance of 8 miles and the neighbourhood Type variability shown in Figure 4.7.

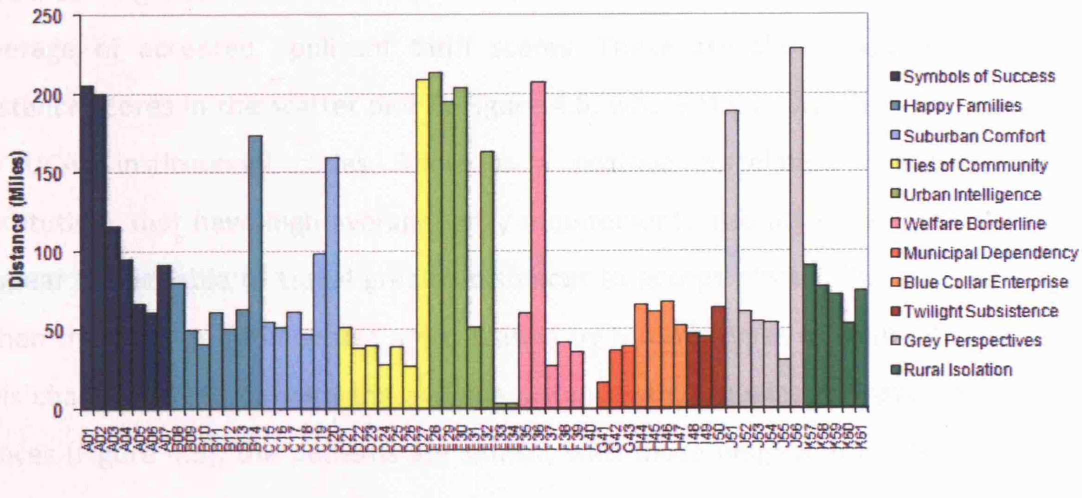


Figure 4.6: Distance Travelled to Accept a Degree Place by Mosaic Types at Lancaster University (Source: 2004 UCAS Data)

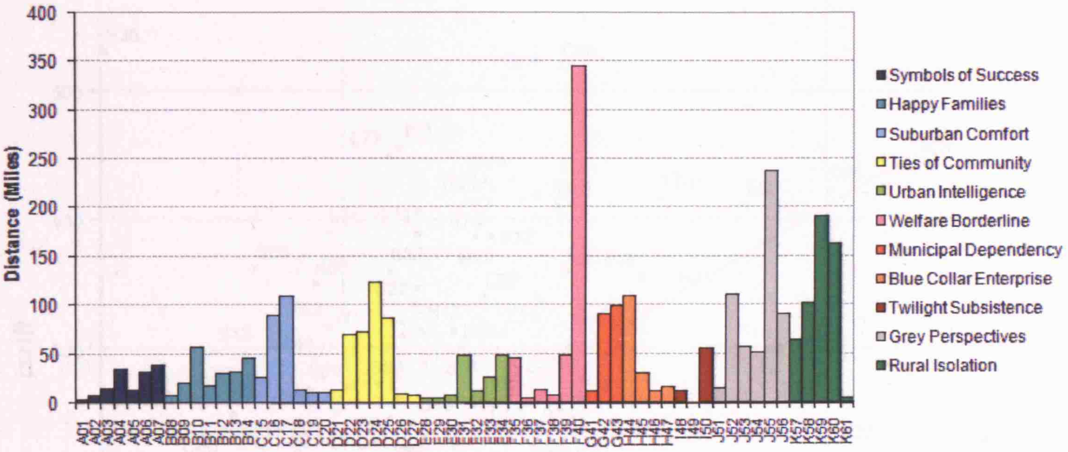


Figure 4.7: Distance Travelled to Accept a Degree Place by Mosaic Types at the University of Westminster (Source: 2004 UCAS Data)

The variability in the overall median distance that is travelled to accept a place is influenced by the proportion of students from different neighbourhood Types that accept places. Therefore, in the example above the University of Westminster will accept a majority of its students from neighbourhood Types with low propensities to travel.

4.3.2 Tariff Scores and Travel Propensities

The incoming attainment level required for each institution may be measured by an average of accepted applicant tariff scores. These are shown against median distance scores in the scatter plot in Figure 4.8, where the point labels correspond to UCAS institutional codes. There is a positive correlation indicating that institutions that have high average entry requirements also attract applicants who appear to feel able to travel greater distances to accept places. This is confirmed when the average tariff scores are classified by Mosaic Types in Figure 4.9. When this chart is compared with the average distances that applicants travel to accept places (Figure 4.5), the patterns are similar, with those neighbourhoods supplying students who travel further also entering with higher attainment.

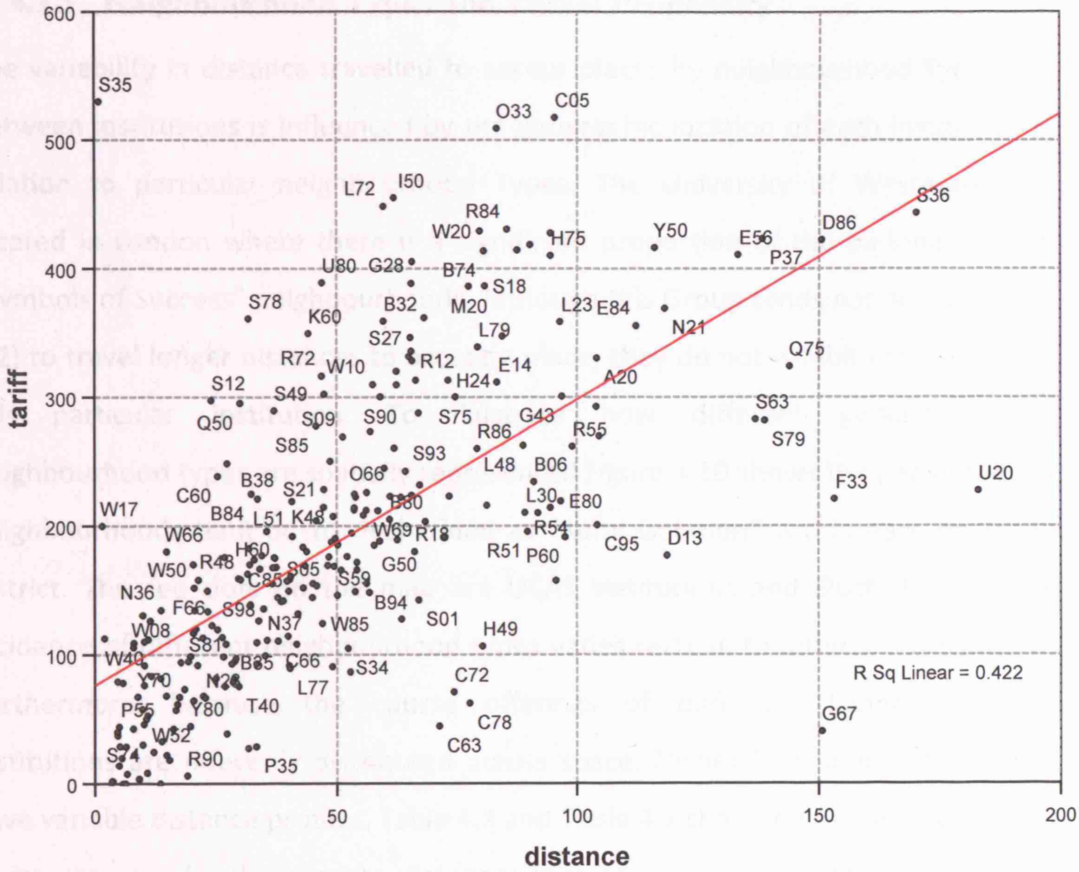


Figure 4.8: A Cross Tabulation of Institutional Average Tariff Scores Against Average Distance Travelled to Accept Degree Offers (Source: 2004 UCAS Data)[28]

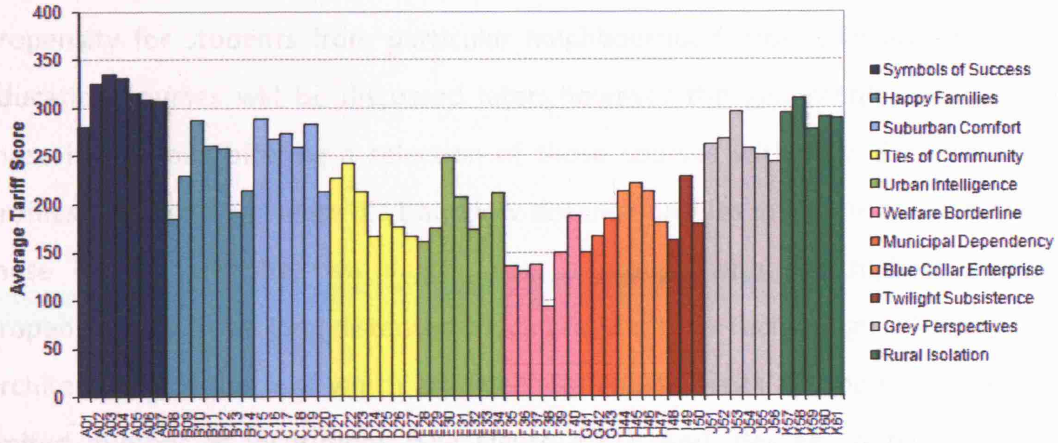


Figure 4.9: Average Tariff Scores by Mosaic Type for Accepted Degree Applicants (Source: 2004 UCAS Data)

[28] For a list of UCAS university codes please see Table 12.2in the Appendix.

4.3.3 Neighbourhood Types and Travel Propensity

The variability in distance travelled to accept places by neighbourhood Types and between institutions is influenced by the geographic location of each institution in relation to particular neighbourhood Types. The University of Westminster is located in London where there is a significant proportion of the national tally of "Symbols of Success" neighbourhoods. Although this Group tends nationally (Figure 3.2) to travel longer distances to accept a place, they do not exhibit this pattern at this particular institution. To illustrate how different geodemographic neighbourhood types are spatially represented Figure 4.10 shows the percentage of neighbourhoods (unit postcodes) coded as "Rural Isolation" within each postcode district. The red dots on the map are UCAS institutions and illustrate how the incidence of different neighbourhood types varies relative to different institutions. Furthermore, because the course offerings of different Higher Education institutions are unevenly distributed across space, Higher Education courses also have variable distance profiles. Table 4.1 and Table 4.2 show the top and bottom 20 degree courses by the average distances that applicants travel to accept places. These measures are partly related to the type of neighbourhoods from which a course typically attracts students, the scarcity of the course nationally and the typical location of the institutions in which the courses are delivered. The propensity for students from particular neighbourhood groups to attend Higher Education courses will be discussed later; however the geographical location of those institutions offering a selection of those courses with high distance travel profiles are shown in Figure 4.11 and low distance profiles in Figure 4.12. Shown in these Figures are the two degree course groups with the highest average propensities to travel long distances which are Maritime Technology (J6) and Naval Architecture (J5), both of which are specialist course areas and being taught at a limited number of institutions (See Figure 4.11); And, degree courses with the lowest propensity to travel (Law by Topic – M2, Historical & Philosophical studies – V0, and Information Systems – G5) have a far wider distribution (See Figure 4.12).

Table 4.1: Top 10 Average Distance Travelled to Accept Degree Courses (Source: 2004 UCAS Data)

Rank	JACS Subject Line	Miles	Acceptances
1	J6 - Maritime Technology	127	153
2	H5 - Naval Architecture	115	113
3	T2 - Japanese studies	113	144
4	T1 - Chinese studies	112	84
5	L6 - Anthropology	110	534
6	T9 - Others in non-European Langs & related	109	897
7	D5 - Forestry	106	61
8	QQ - Combinations within Linguistics, Classics & related	105	498
9	D1 - Pre-clinical Veterinary Medicine	104	873
10	V3 - History by Topic	104	1113
11	Q8 - Classical studies	103	800
12	VV - Combinations within Hist & Philosophical studies	98	1188
13	F7 - Ocean Sciences	98	214
14	C3 – Zoology	97	1105
15	C2 – Botany	96	23
16	V5 – Philosophy	96	1327
17	V6 - Theology and Religious studies	94	1156
18	RR - Combinations within European Langs, Lit and related	92	1838
19	LL - Combinations within Social Studies	92	2436
20	F6 - Geology	91	1344

Table 4.2: Bottom 10 Average Distance Travelled to Accept Degree Courses (Source: 2004 UCAS Data)

Rank	JACS Subject Line	Miles	Acceptances
1	M2 - Law by Topic	32	1263
2	V0 - Hist & Philosophical studies: any area	32	284
3	G5 - Information Systems	32	3753
4	BB – Combinations within Subjects allied to Medicine	36	223
5	L5 - Social Work	37	5330
6	XX – Combinations within Education	37	227
7	X3 - Academic studies in Education	37	3784
8	X1 - Training Teachers	38	7636
9	KK – Combinations within Architecture, Build & Plan	38	163
10	F0 - Physical Sciences: any area of study	39	215
11	W9 - Others in Creative Arts and Design	39	539
12	MM - Combinations within Law	39	259
13	X9 - Others in Education	40	163
14	B7 – Nursing	40	5684
15	P4 – Publishing	42	176
16	B8 - Medical Technology	42	1669
17	L4 - Social Policy	43	856
18	N4 – Accounting	44	5846
19	G6 - Software Engineering	45	1259
20	N6 - Human Resource Management	46	768

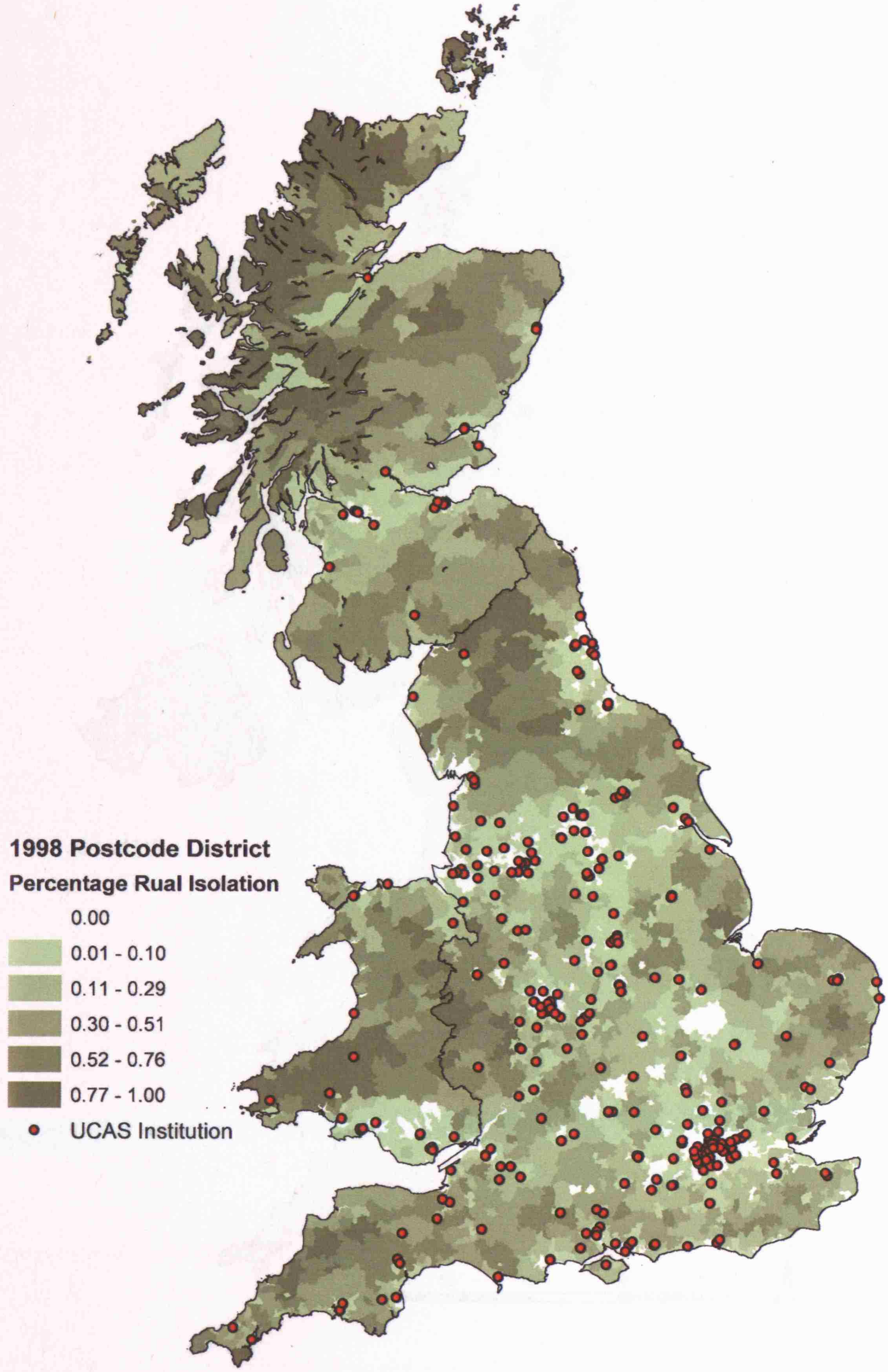


Figure 4.10: Percentage of Mosaic "Rural Isolation" Postcodes (Source: Experian Mosaic and Author Calculations)

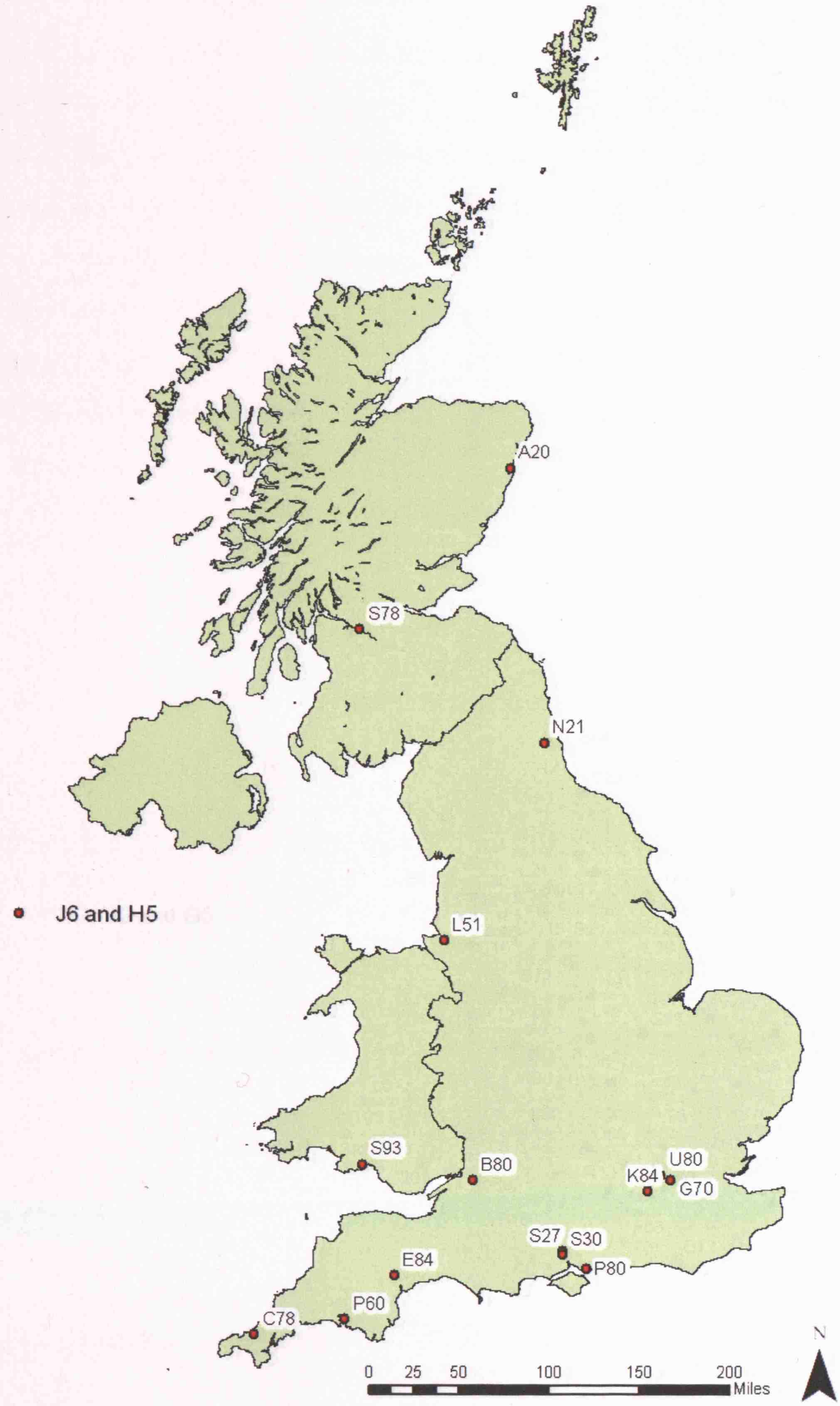


Figure 4.11: Institutions with Degree Acceptances on Courses in Maritime Technology -J6 and Naval Architecture -J5 (Source: 2004 UCAS Data)

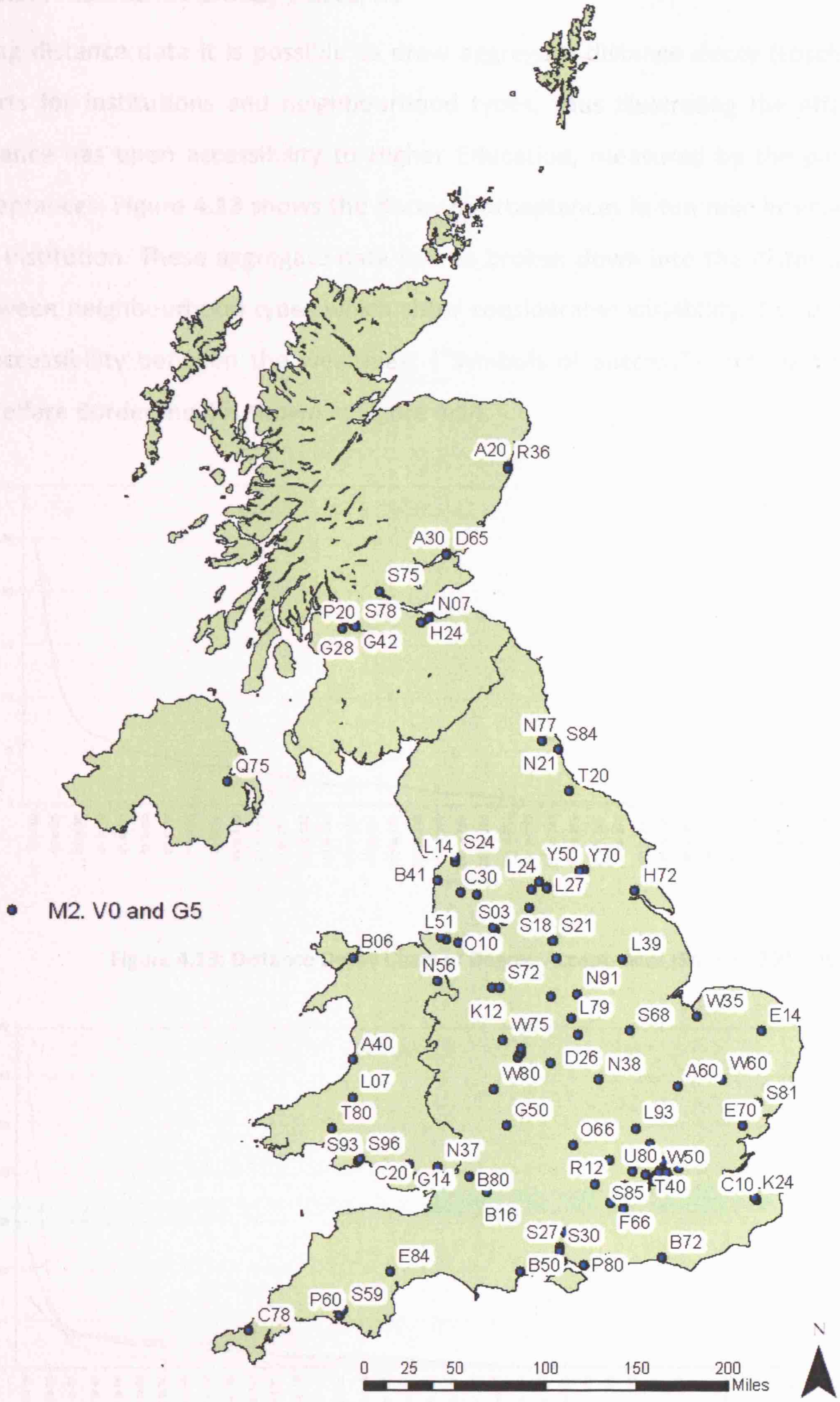


Figure 4.12: Institutions with Degree Acceptances on Courses in Law by Topic – M2, Historical & Philosophical Studies – V0, and Information Systems – G5 (Source: 2004 UCAS Data)

4.3.4 Distance Decay Patterns

Using distance data it is possible to draw aggregate distance decay (Lösch, 1954) charts for institutions and neighbourhood types, thus illustrating the effect that distance has upon accessibility to Higher Education, measured by the pattern of acceptances. Figure 4.13 shows the decay in acceptances in ten mile intervals from the institution. These aggregate data can be broken down into the distance decay between neighbourhood types which show considerable variability. The difference in accessibility between the wealthiest ("Symbols of Success") and least wealthy ("Welfare Borderline") is shown in Figure 4.14.

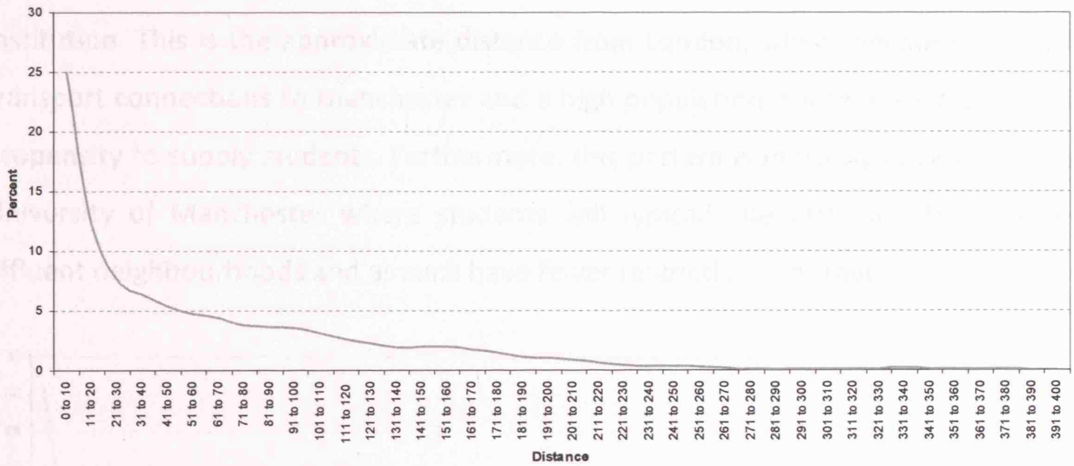


Figure 4.13: Distance Decay Chart of Degree Acceptances (Source: 2004 UCAS Data)

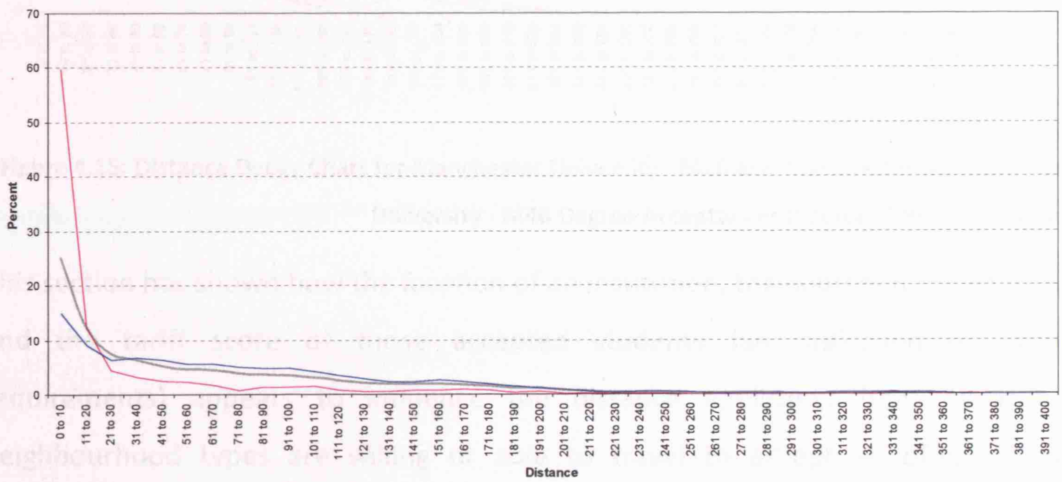


Figure 4.14: Distance Decay Chart of "Welfare Borderline" and "Symbols of Success" Degree Acceptances (Source: 2004 UCAS Data)

Around 60% of all acceptances from "Welfare Borderline" (pink line) neighbourhoods come from within 20 miles of an institution and compares with just 15% from "Symbols of Success" (blue line) neighbourhoods. The mix of neighbourhood types an institution attracts creates similar aggregate distance decay profiles. The institutions "M40" Manchester Metropolitan University and "M20" the University of Manchester, both of which are in the same city and within 1 mile of each other have the distance decay profiles which can be found in Figure 4.15. Manchester Metropolitan University (M40) attracts 15% more of its applicants than Manchester University from within 10 miles of the institutions. Of further note is the latter peak visible in both universities at around 160 miles from the institution. This is the approximate distance from London, where because of good transport connections to Manchester and a high population, there is an increase in propensity to supply students. Furthermore, this pattern is more significant for the University of Manchester where students will typically be attracted from more affluent neighbourhoods and as such have fewer restrictions on travel.

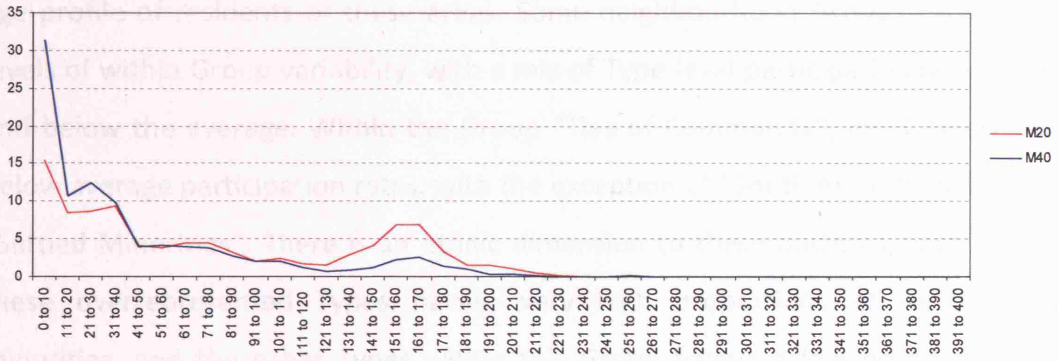


Figure 4.15: Distance Decay Chart for Manchester University - M20 and Manchester Metropolitan University - M40 Degree Acceptances (Source: 2004 UCAS Data)

This section has shown how the location of an institution, the courses it has to offer and the tariff score of those accepted students (an indication of entry requirements) appears to influence the distance applicants from particular neighbourhood types are willing or able to travel to accept an offer. These interacting spatial complexities affect both individual course and institutional market areas, and have implications for the targeting of particular student groups for purposes of marketing or widening participation.

4.4 Geodemographics and Social-Economics

The previous section has shown the apparent discontinuities in travel to accept places at Higher Education, both spatially and by neighbourhood type. To further elaborate how variation in Higher Education access geographies are manifested a series of charts based upon index scores are discussed in the following section. These illustrate how neighbourhood typologies can provide an effective tool to analyse socio-spatial differences in Higher Education data. An aggregate profile for 2004 home Higher Education acceptances can be created by comparing the proportion of all UK degree acceptances by neighbourhood type with total adult population (see Figure 4.16). This shows how 2004 degree acceptance rates vary according to neighbourhood Types. There are large between Group differences, with "Symbols of Success" showing the greatest propensity to participate, and those from "Welfare Borderline", "Municipal Dependency" and "Blue Collar Enterprise" having the lowest. "Twilight Subsistence" and "Grey Perspectives" also have a low propensity to participate, although this is likely to be caused by the older age profile of residents of these areas. Some neighbourhood Groups exhibit high levels of within Group variability, with a mix of Type level participation rates above and below the average. Within the Group "Ties of Community", most Types have below average participation rates, with the exception of "South Asian Industry" and "Settled Minorities". There is an ethnic dimension to these patterns, with both of these overrepresented Types having very high proportions of Asian ethnic minorities, and the other Types within this Group having a low proportion. The importance placed on Higher Education by these neighbourhood Types could explain these differences. Similarly, within the Group "Welfare Borderline", the Type "Metro Multiculture" which defines neighbourhoods that typically have a high proportion of Asian residents correspondingly shows a high propensity to participate in Higher Education.

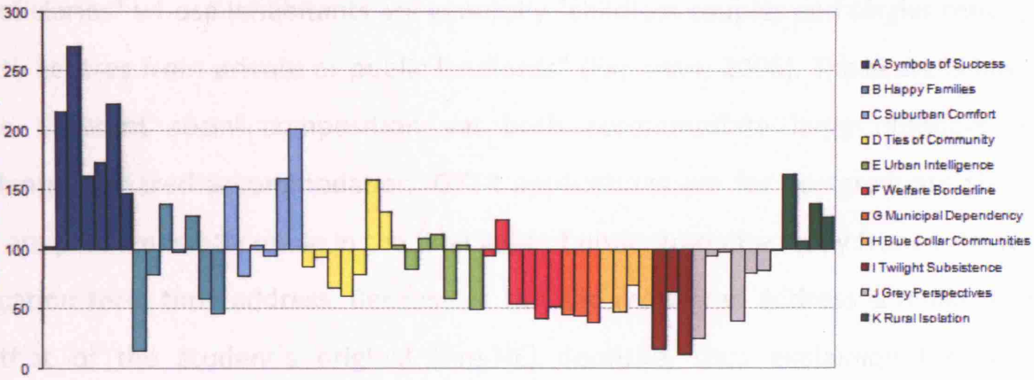


Figure 4.16: Degree Acceptances by Neighbourhood Type (Source: 2004 UCAS Data)

In order illustrate a series of characteristics about the use of neighbourhood classifications within Higher Education an alternative admissions system is presented in the following analysis. To qualify as a Primary or Secondary School Teacher a qualification recognised by the Graduate Teacher Training Registry (GTTR) must be obtained. The GTTR administers applications to these courses through UCAS. The neighbourhood profiles for 2004 acceptances to these courses are shown in Figure 4.17.

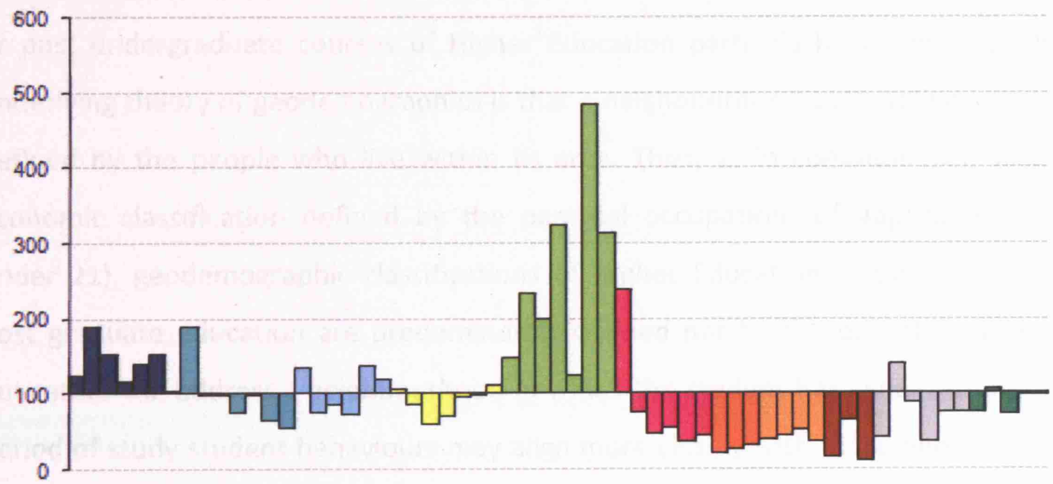


Figure 4.17: Propensity for 2004 GTTR Acceptances by Neighbourhood Type (Source: 2004 UCAS Data)

There is a high propensity for acceptances from the group "Urban Intelligence", which are areas with people who are "young, single and mostly well-educated, and who are cosmopolitan in tastes and liberal in attitudes" (Experian, 2006). A further Type which stands out from the Group "Welfare Borderline" is "Bedsit

Beneficiaries" whose inhabitants are generally "childless couples and singles renting in city centres from private or public landlords" (Experian, 2006). These areas have quite different social composition yet both accommodate large numbers of students in shared accommodation. GTTR applications are for postgraduate study and are predominantly made in the final year of undergraduate study from a Higher Education term time address. Because of this the registered address is unlikely to be that of the student's original (Pre-HE) domicile, thus explaining the high incidence of these Types. This also illustrates two further points about the use of geodemographics for Higher Education applications. When an application is made to UCAS, applicants are asked for both a "home" address and a "contact" address. The addresses used throughout this thesis are for "home" addresses, which are characteristic of the type of neighbourhood or background that a student has come from. There are inherent and undetectable errors with this process as a student may apply for a new course of Higher Education while living away at university. It is possible that a student may specify the same "contact" and "home" address, as they perhaps view their current accommodation as their home. The GTTR profile also demonstrates a further issue and one which makes targeting students during or post undergraduate courses of Higher Education particularly challenging. The underlying theory of geodemographics is that a neighbourhood both defines, and is defined by the people who live within its area. Thus, as in common with socio-economic classification defined by the parental occupation (of applicants aged under 21), geodemographic classifications of Higher Education in the context of post graduate education are predominantly defined not by the parental address, but rather the address / neighbourhood in which the student has lived. During the period of study student behaviours may align more closely with those people living within the new neighbourhood containing the student accommodation. It is reasonable to anticipate change in student behaviour when living away from home, e.g. with respect to the propensity to purchase beer. It is for this reason that the GTTR profile is defensible, as by the end of an undergraduate programme it is likely that actual student behaviour will be more homogenous than the possibly more diverse range of neighbourhood types defined according to parental home address.

The further scheme profile that is shown in Figure 4.18 is derived from the Nursing and Midwifery Admissions System (NMAS) which is also managed by UCAS. NMAS processes admissions to diploma and degree courses which lead to qualified nurse status. There is an overrepresentation of the neighbourhood Groups "Happy Families" which are areas where "families focus on career and home, are mostly younger age groups now raising children" (Experian, 2006), and "Ties of Community" which contain "people living in close-knit inner city and manufacturing town communities, responsible workers with unsophisticated tastes" (Experian, 2006). The former of these neighbourhood Groups is considerably more affluent than the latter.

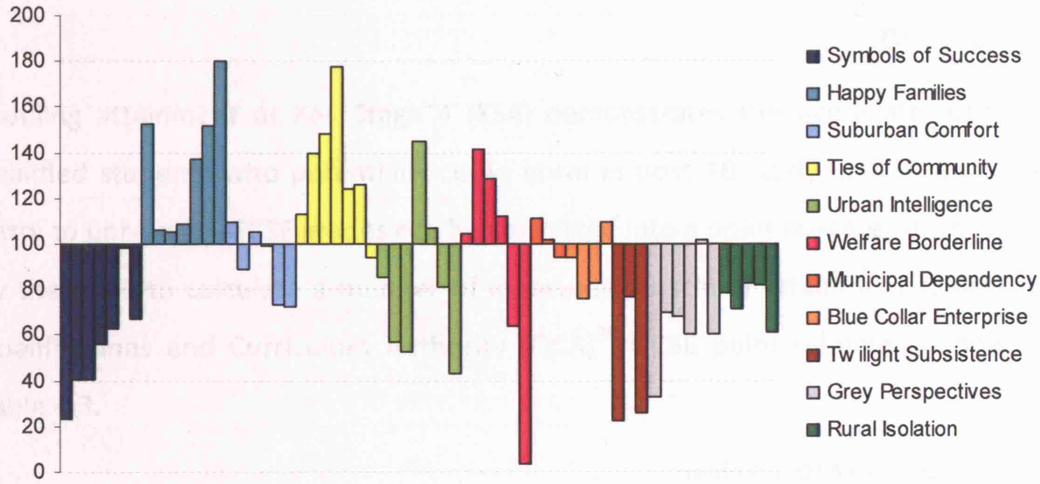


Figure 4.18: NMAS Acceptances by Neighbourhood Type (Source: 2004 UCAS Data)

4.4.1 Segmentation by Prior Attainment

Access to Higher Education requires an applicant to have prior attainment at A Level or a nationally-recognised equivalent, in order to fulfil entry criteria that may stipulate a combination of grades and qualification types. Differences between students in terms of attainment therefore largely determine opportunities to access certain Higher Education courses or institutions and this, along with the demand for courses and subject profile, in turn affects an institution's aggregate profile of incoming students. For example, an applicant for a degree in Medicine would be expected to possess A-Levels in science, and without these qualifications would not be successful in gaining access. These issues have been addressed in the literature by Reay *et al* (2005:87) who argue that "choice in Higher Education is constrained by the predicted and actual grades achieved by the students". These patterns are further entrenched by requirements for specialism in prior qualification which are deemed essential by subject specialists in Higher Education institutions. These effects were recognised in the 2003 White Paper on Higher Education:

> ***"The problem does not begin at age 16. Recent research suggests that a significant difference appears even before children have reached the age of two years. The attainment gap continues to widen through the phases of education, although the pace of increase slows down once children reach 7 years of age. Analysis suggests that at least three quarters of the 30 percentage point social gap in higher education participation can be attributed to differences in the level of attainment by the age of 16. Thirty per cent of children whose parents are in unskilled occupations achieve five or more good GCSEs, compared to 69 per cent of children whose parents are professional or managerial."***

(DfES, 2003: 7)

Profiling attainment at Key Stage 4 (KS4) demonstrates the aggregate profile of qualified students who potentially could enrol in post 16 study, and as such gain entry to university. GCSE grades can be converted into a point scheme which is used by the DCSF to calculate a number of measures on school attainment tables. The Qualifications and Curriculum Authority (QCA)[29] GCSE point scheme is shown in Table 4.3.

Table 4.3: QCA GCSE Point Scheme

Grade	Points
A*	58
A	52
B	46
C	40
D	34
E	28
F	22
G	16

Using this point system an average score across neighbourhood Groups has been calculated (see Figure 4.19). There are clear differences between the

[29] http://www.qca.org.uk/

neighbourhood Groups which can be examined further with reference to GCSE grade profiles (see Table 4.4).

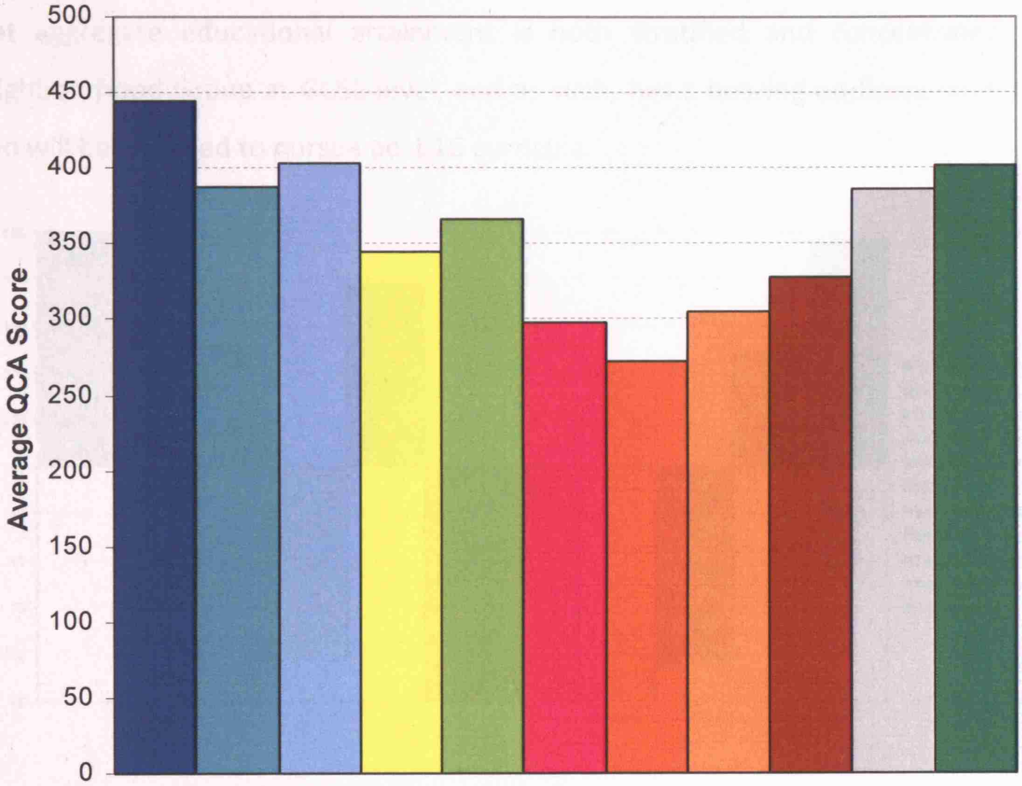


Figure 4.19: Average KS4 2006 GCSE Points by Mosaic Groups (Source: DCSF)

Table 4.4: Index Scores for 2006 KS4 GCSE Grades by Mosaic Groups (Source: DCSF)

Grade	A*	A	B	C	D	E	F	G
Symbols of Success	234	179	133	91	63	43	31	22
Happy Families	93	105	110	108	99	88	75	63
Suburban Comfort	126	125	118	104	89	75	61	51
Ties of Community	59	74	89	104	112	117	122	124
Urban Intelligence	139	119	105	93	93	89	89	95
Welfare Borderline	45	56	73	92	118	143	168	191
Municipal Dependency	17	32	53	88	125	168	213	262
Blue Collar Enterprise	36	51	70	97	123	147	164	167
Twilight Subsistence	58	69	85	102	114	124	129	133
Grey Perspectives	124	120	112	102	93	81	71	61
Rural Isolation	149	135	119	101	86	70	54	44

Table 4.4 is visually represented in Figure 4.20 - Figure 4.27 which demonstrates neighbourhood inequality in attainment across grade boundaries. Those who are living in the most affluent areas achieve the highest grades. These data clearly show that aggregate educational attainment is both stratified and concentrated by neighbourhood Group at GCSE level, and as such, has a bearing on those students who will be qualified to pursue post 16 curricula.

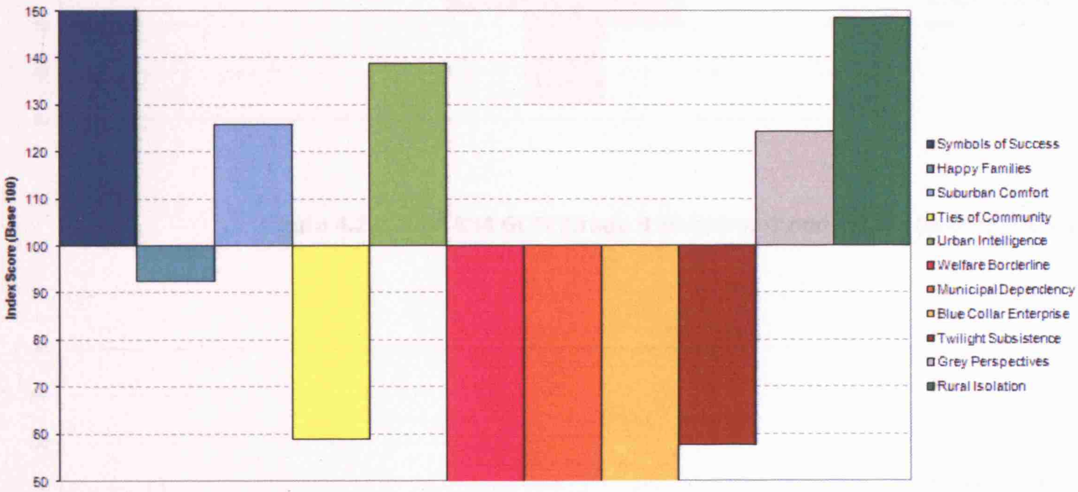


Figure 4.20: 2006 KS4 GCSE Grade A* Neighbourhood Profile (Source: DCSF)

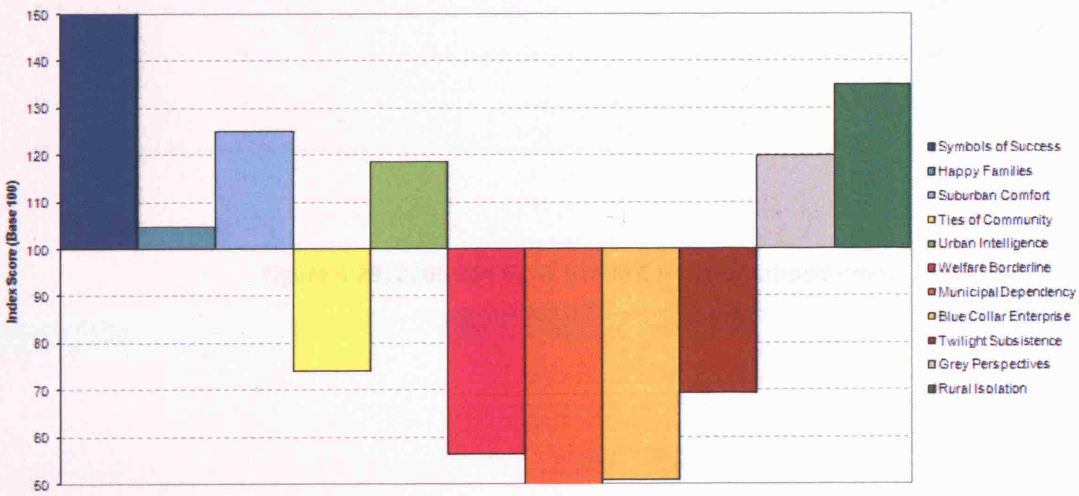


Figure 4.21: 2006 KS4 GCSE Grade A Neighbourhood Profile (Source: DCSF)

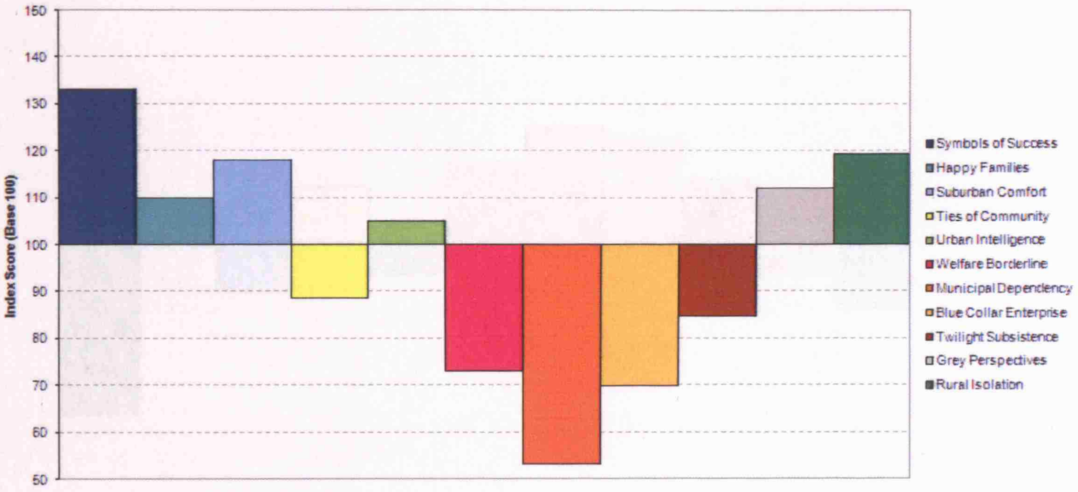


Figure 4.22: 2006 KS4 GCSE Grade B Neighbourhood Profile (Source: DCSF)

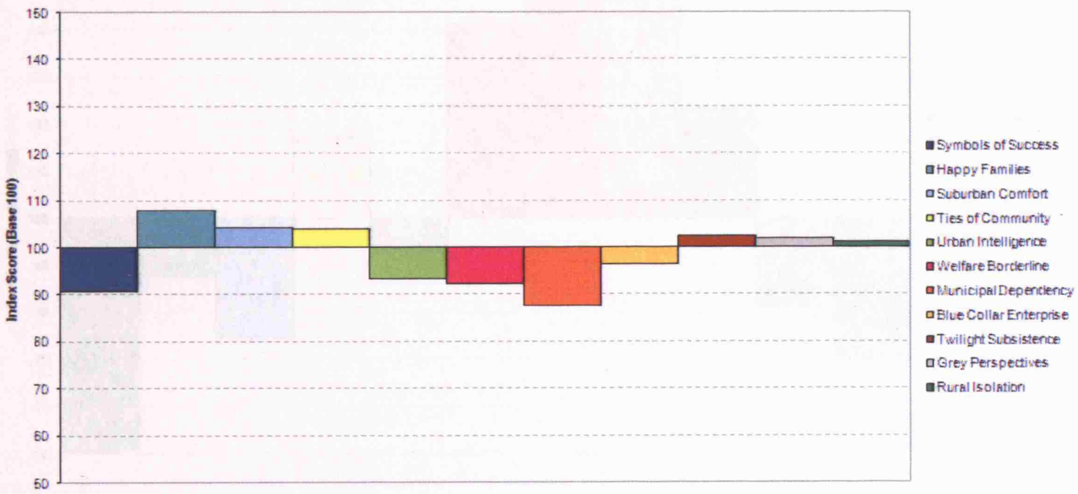


Figure 4.23: 2006 KS4 GCSE Grade C Neighbourhood Profile (Source: DCSF)

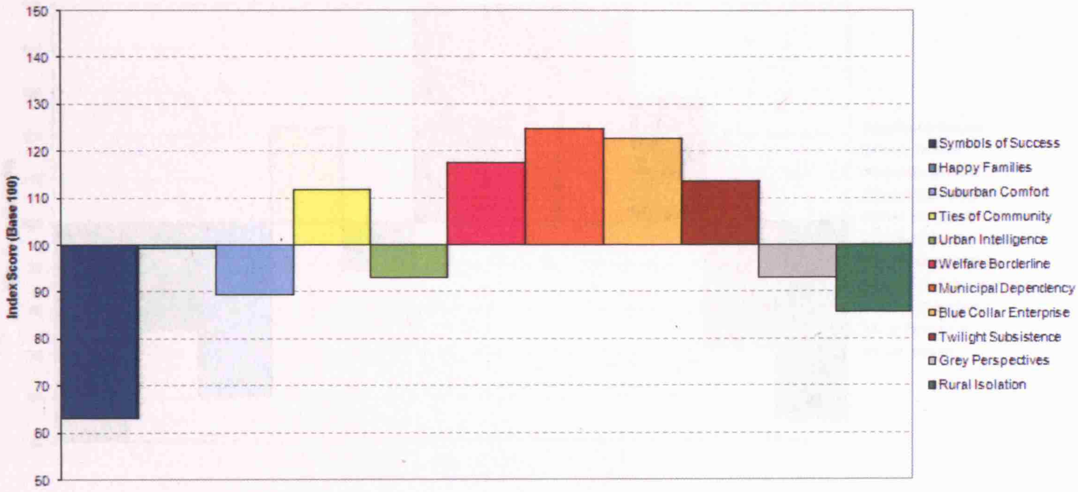


Figure 4.24: 2006 KS4 GCSE Grade D Neighbourhood Profile (Source: DCSF)

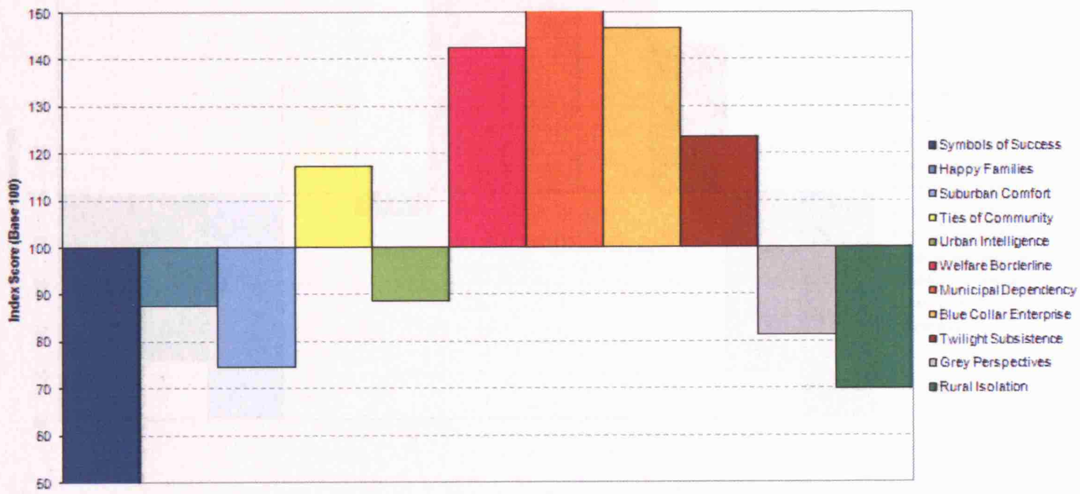


Figure 4.25: 2006 KS4 GCSE Grade E Neighbourhood Profile (Source: DCSF)

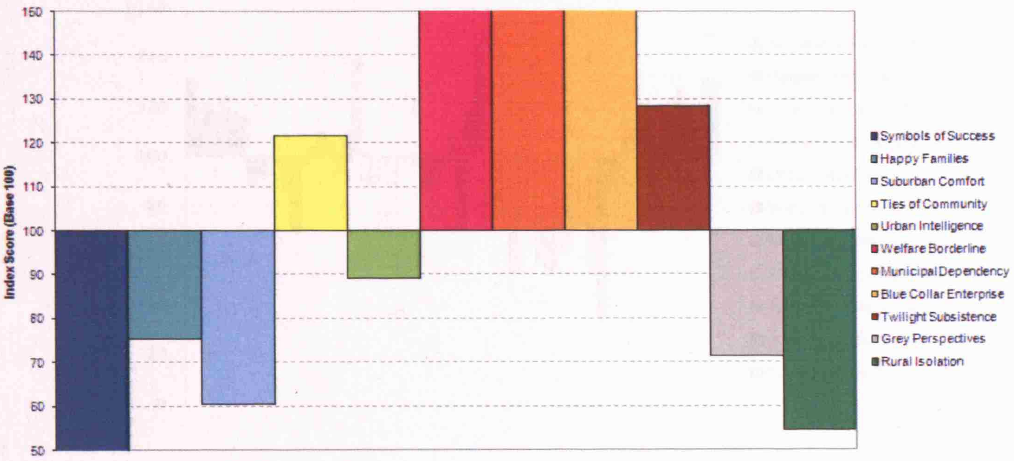


Figure 4.26: 2006 KS4 GCSE Grade F Neighbourhood Profile (Source: DCSF)

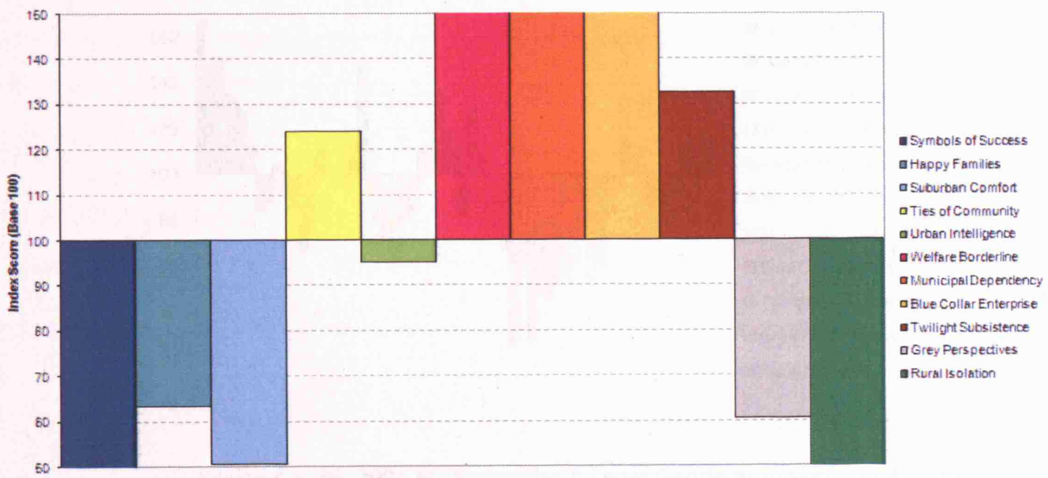


Figure 4.27: 2006 KS4 GCSE Grade G Neighbourhood Profile (Source: DCSF)

The previous analysis has shown how inequality exists between aggregate neighbourhood GCSE scores and the following section will show how similar patterns occur at A-Level. Before a student can attain grade in a subject at A-Level they first must gain a place on a course at a school, college or other tertiary sector institution. These rates of access are variable which provides a first filter on those qualifying students who may be available to study courses of Higher Education. Thus, using the 2006 Key Stage 5 (KS5) PLASC data Figure 4.28 - Figure 4.31 present a series of geodemographic profiles for the areas in which students live who are studying particular A-Level subjects.

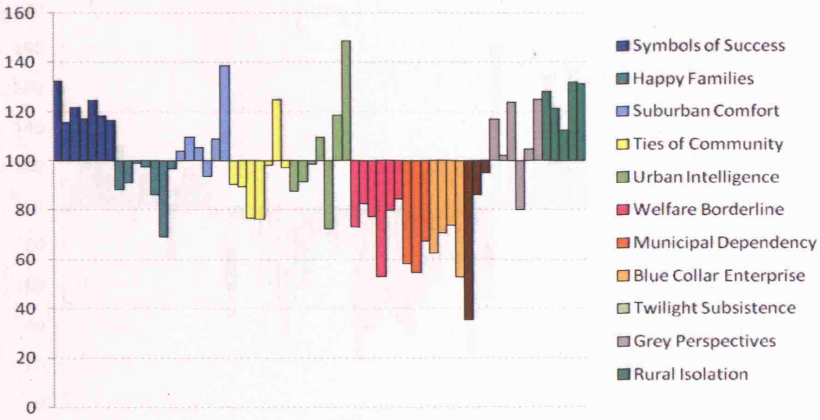


Figure 4.28: 2006 Biology A-Level Geodemographic Profile (Source: DCSF)

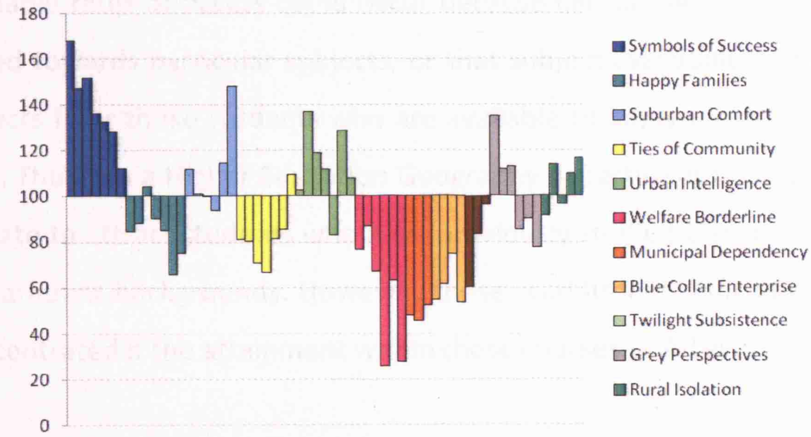


Figure 4.4.29: 2006 Mathematics A-Level Geodemographic Profile (Source: DCSF)

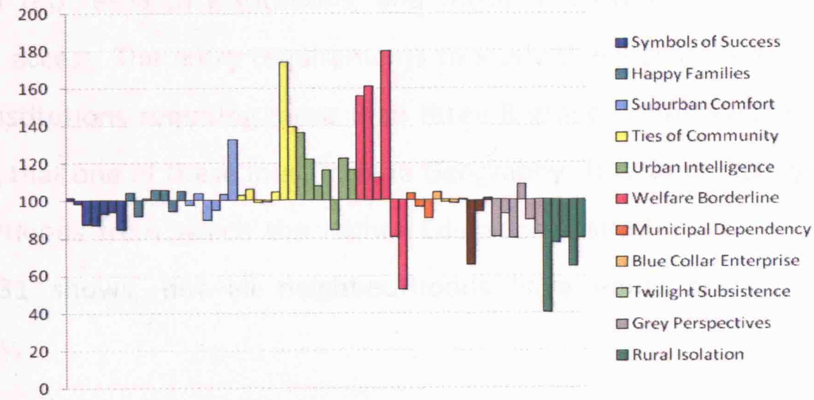


Figure 4.30: 2006 Sociology A-Level Geodemographic Profile (Source: DCSF)

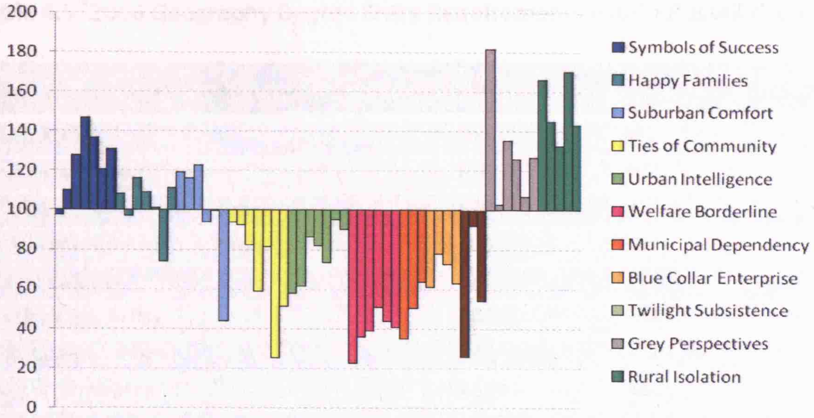


Figure 4.31: Geography A-Level Geodemographic Profile (Source: DCSF)

These variable rates of access could occur because certain neighbourhood Groups are inclined towards particular subjects, or that subject availability varies spatially. These effects filter those students who are available to apply for courses in Higher Education. Thus, for a Higher Education Geography department it becomes difficult on aggregate to attract students who have previously studied Geography at A-Level from less affluent backgrounds. However, these recruitment difficulties are shown to be concentrated if the attainment within these courses at A-Level are examined.

Table 4.5 shows the admissions criteria to Geography for the Russell Group as a sample of top research institutions, and those that often are criticised for not extending access. The entry requirements to study these courses are all very high with all institutions requiring more than three B grades at A-Level and many also specifying that one of these must include Geography. This immediately limits those neighbourhoods from which the Higher Education institution can recruit, and as Figure 4.31 shows, not all neighbourhoods have equal propensity to study Geography.

Table 4.5: 2006 Geography Degree Entry Requirements within Russell Group Institutions

University	Entry Requirement	Geography Required?
Cardiff University	300 pts*	Yes
Imperial College London	N/A	N/A
King's College London	BBB	Yes
London School of Economics & Political Science	BBB	No
Newcastle University	ABB	Yes
Queen's University Belfast	BBC-BCCb**	Yes
University College London	AABe-ABBe	Yes
University of Birmingham	ABB-BBB	Yes
University of Bristol	AAA-AAB	Yes (not for BSc)
University of Cambridge	AAA	No
University of Edinburgh	BBB	Yes
University of Glasgow	ABB	No
University of Leeds	ABB	Yes
University of Liverpool	320 pts	No
University of Manchester	AAB-ABB	No
University of Nottingham	AAA-AAB	Yes
University of Oxford	AAA-AAB	Yes
University of Sheffield	ABB-ABbb	Yes
University of Southampton	AAB-ABB	Yes
University of Warwick	N/A	N/A

* Minimum points required from qualifications with the volume and depth of A level or equivalent.
**Lowercase letters indicate an AS-Level

Figure 4.32 to Figure 4.36 demonstrate how these issues are compounded if attainment within Geography A-Levels is profiled. For institutions wishing to recruit "A" grade Geography A-Level students there is bias towards students achieving these grades to live in more affluent areas. This makes it very difficult for Russell Group institutions to extend their offers into those areas which typically would not participate in Higher Education.

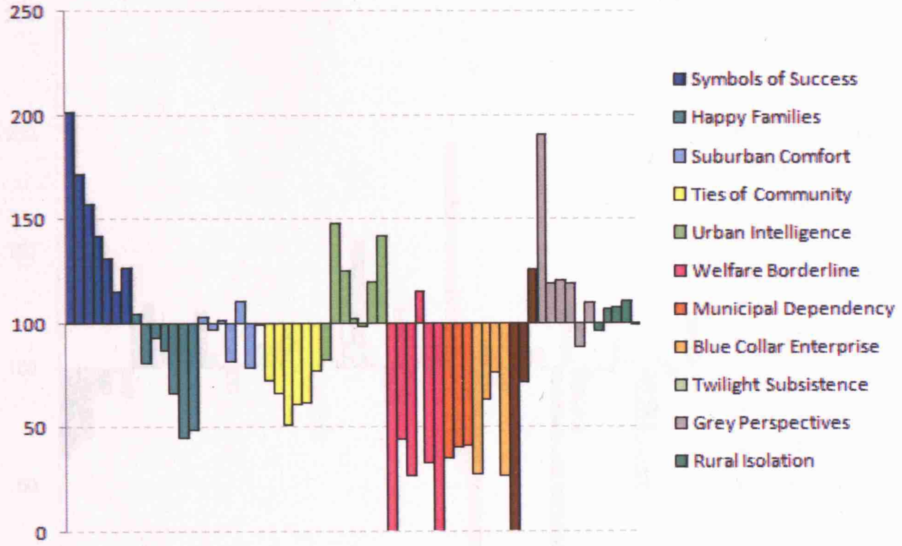


Figure 4.32: "A" at A-Level Geography in 2006 (Source: DCSF)

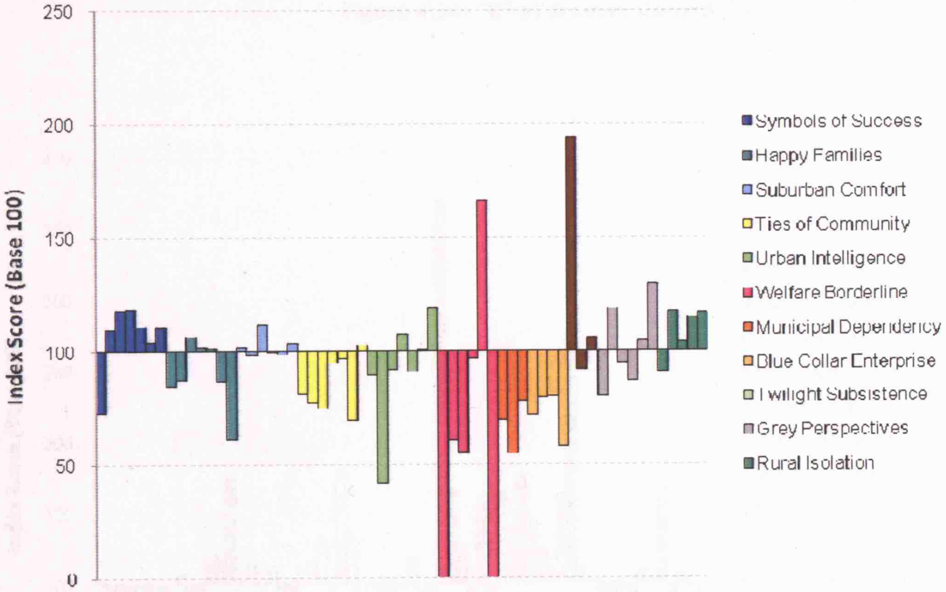


Figure 4.33: "B" at A-Level Geography in 2006 (Source: DCSF)

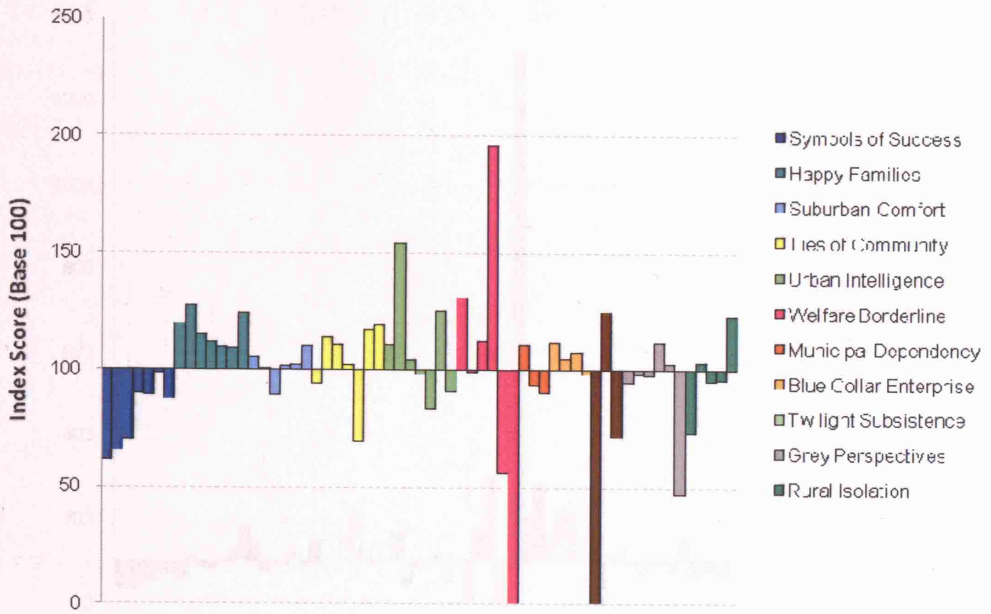


Figure 4.34: "C" at A-Level Geography in 2006 (Source: DCSF)

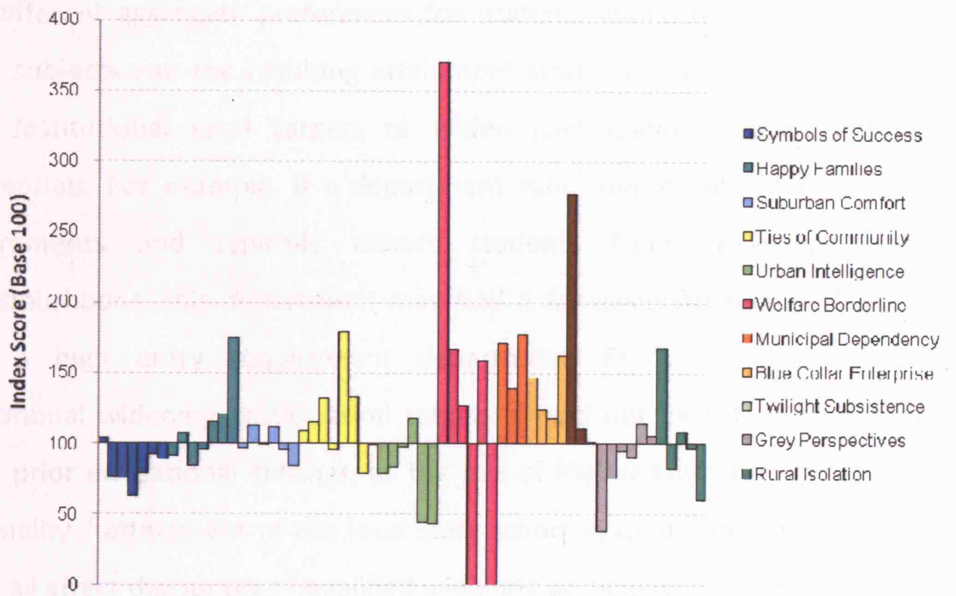


Figure 4.35: "D" at A-Level Geography in 2006 (Source: DCSF)

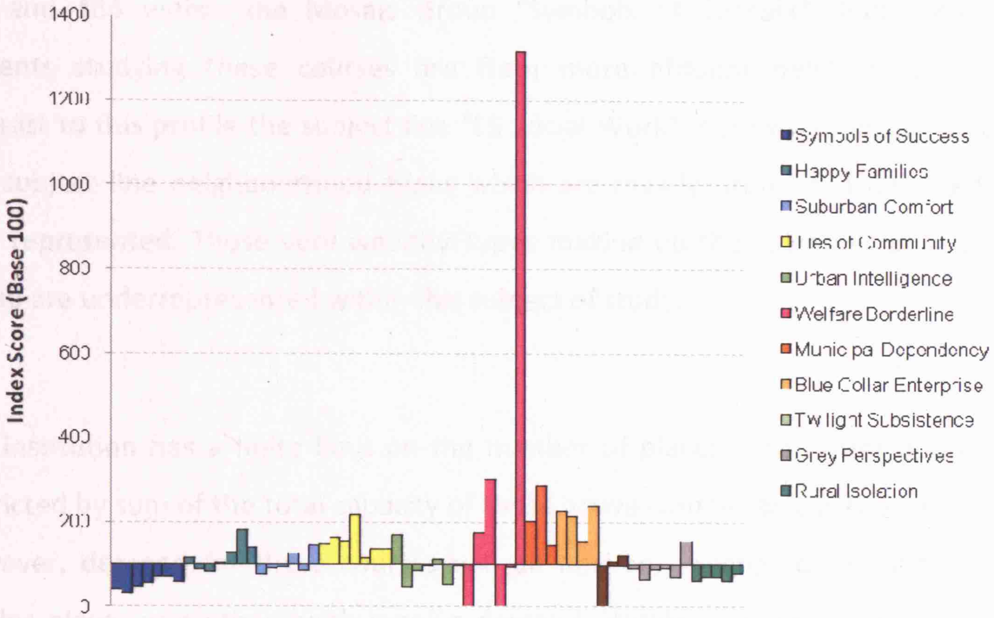


Figure 4.36: "E" at A-Level Geography in 2006 (Source: DCSF)

The different aggregate preferences for students within neighbourhood Types to study subjects and the resulting attainment stratification mean that national and even institutional level targets to widen participation should address these differentials. For example, if a department runs courses which have lower entry requirements and typically attract students from widening participation neighbourhoods, this department may find it far easier to recruit these students than a high entry requirement department. From a national perspective institutional widening participation targets should not be set without considering these prior educational settings, as the mix of Higher Education courses on offer, the quality / attainment of the local state school system and the courses they run could all affect the supply of qualified widening participation students.

4.4.2 Course Choice and Different Neighbourhood Origins

Index scores are calculated for a number JACS Course lines in Figure 4.37 and Figure 4.38. These show the propensity for participation from different neighbourhood Mosaic Types. If a course had a geodemographic profile in exactly the same proportions as the average Higher Education profile, all scores would be 100. For courses in subject line "A1 Pre-Clinical Medicine", Type index scores range between

120 and 184 within the Mosaic Group "Symbols of Success", indicating that students studying these courses are from more affluent neighbourhoods. In contrast to this profile the subject line "L5 Social Work" is shown in Figure 4.38. In this subject line neighbourhood types which are mainly urban and deprived are over represented. Those very wealthy Types making up the "Symbols of Success" Group are underrepresented within this subject of study.

Any institution has a finite limit on the number of places it can offer applicants, restricted by sum of the total capacity of those active courses across the institution. However, demand for these courses will be uneven, leaving some courses with surplus places and others with surplus demand. Within a constrained number of course places geodemographic analysis could provide two useful tools to an institution to maximise the efficiency of the offer process. An applicant applying to an oversubscribed course could potentially be offered a place on a course that is in less demand, but informed by the information that there is a high probability that applicants form this neighbourhood Type would accept this offer. For example, courses in Nutrition may be oversubscribed, although surplus applicants might be redirected towards available places in Anthropology rather than rejecting them outright as there is a high propensity for acceptances to both these subject lines to be from the neighbourhood group "Urban Intelligence". More strategically, widening participation bursaries could be distributed to those courses which have the greatest under representation of those neighbourhood types typically underrepresented in Higher Education. If, for example, a course had 30 fewer students from these neighbourhood groups than would be expected, it may be that 30 bursaries should be provided to encourage these students to participate. However, provision of such bursaries does not guarantee that applicants will apply or indeed accept offers, and as such the rate of return on bursaries available would have to be factored into the number offered, e.g. 50 may need to be offered to get a return of 30 acceptances. Furthermore, such initiatives could be described as attempts at social engineering, and any attempt to alter course profiles will reduce the probability of non targeted students being allocated or accepting places on these courses. A decision may therefore be taken to target only those courses that have surplus capacity. This raises the important issue of whether access to Higher Education can be "fairly" extended to underrepresented groups in institutions where capacity remains fixed and oversubscribed. One view would simply be that the best qualified students alone should be offered places, yet because of the inequalities in the level (and in some cases type) of incoming attainments from widening participation groups, this would always result in them being under-represented. A method of accounting for social inequalities in the offer making process is necessary in order to achieve a socially broader allocation of places across both over- and under- subscribed courses.

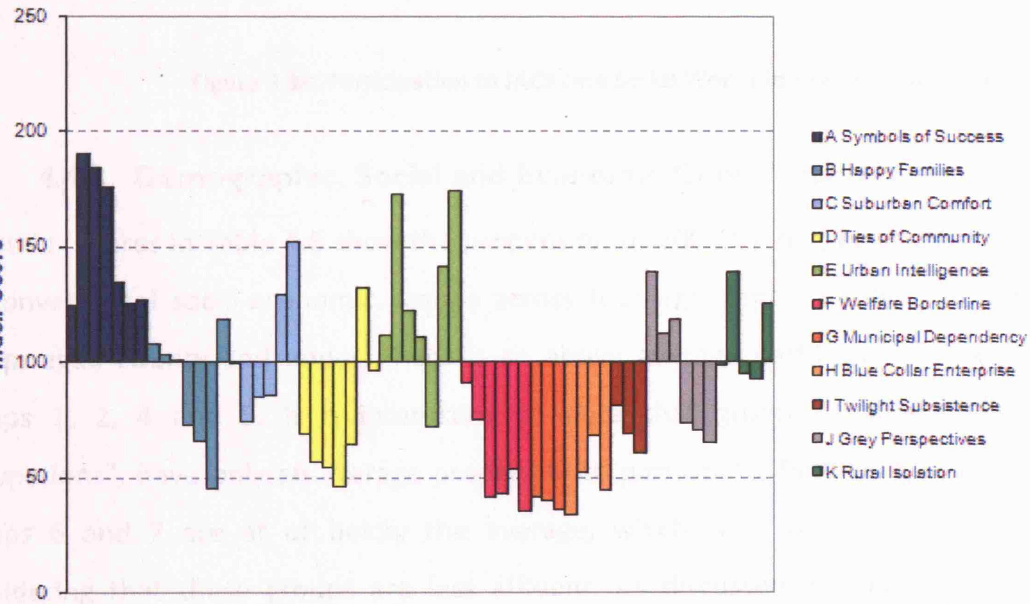


Figure 4.37: Participation to JACS Line Pre-clinical Medicine (Source: 2004 UCAS Data)

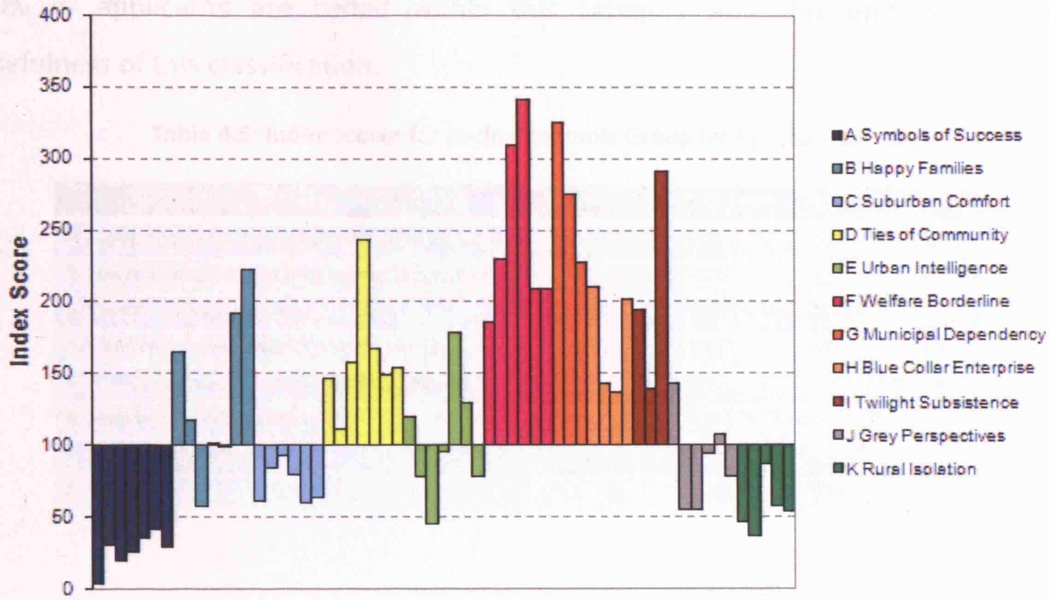


Figure 4.38: Participation to JACS Line Social Work (Source: 2004 UCAS Data)

4.4.3 Demographic, Social and Economic Considerations

The index scores in Table 4.6 show the propensity for 2004 UK degree acceptances by conventional socio-economic groups across four age bands. In the youngest group, aged twenty and under, there is an above average participation rate for groups 1, 2, 4 and 5. It is interesting to note that group 3, "Intermediate Occupations", have only an average propensity to participate. Participation across groups 6 and 7 are at or below the average, which is as one would expect considering that these groups are less affluent. As discussed in Chapter 2 this principal student age cohort is assigned socio-economic group based on the parental occupations, which not only requires the applicant to know what the job the parent does, but also for the short description of this job to be correctly coded into a classification. After the age of 21, the classification is based on the occupation of the applicant, and so the data are not directly comparable. This likely accounts for the significant decrease in the size of groups 1 and 2 across the older age cohorts. There is also increased propensity in the three older groups for the applicant to be coded as "unknown", thus indicating that there may be some systematic error in the conversion of occupation to socio-economic group. Overall, 20% of applicants are coded within this category and this undermines the usefulness of this classification.

Table 4.6: Index Scores for Socio-Economic Group by Age (Source: 2004 UCAS Data)

	20 and under	21 to 24	25 to 39	40 +
1. Higher managerial and professional occupations	117	28	29	39
2. Lower managerial and professional occupations	109	53	67	85
3. Intermediate occupations	98	96	123	116
4. Small employers and own account workers	113	49	47	48
5. Lower supervisory and technical occupations	114	45	38	44
6. Semi-routine occupations	90	144	145	121
7. Routine occupations	100	124	88	55
8. Unknown	71	233	207	196

Figure 4.39 shows the neighbourhood profile for those student acceptances aged over 19 years. This chart demonstrates how mature students more predominantly come from neighbourhoods which are less affluent when compared to young participants.

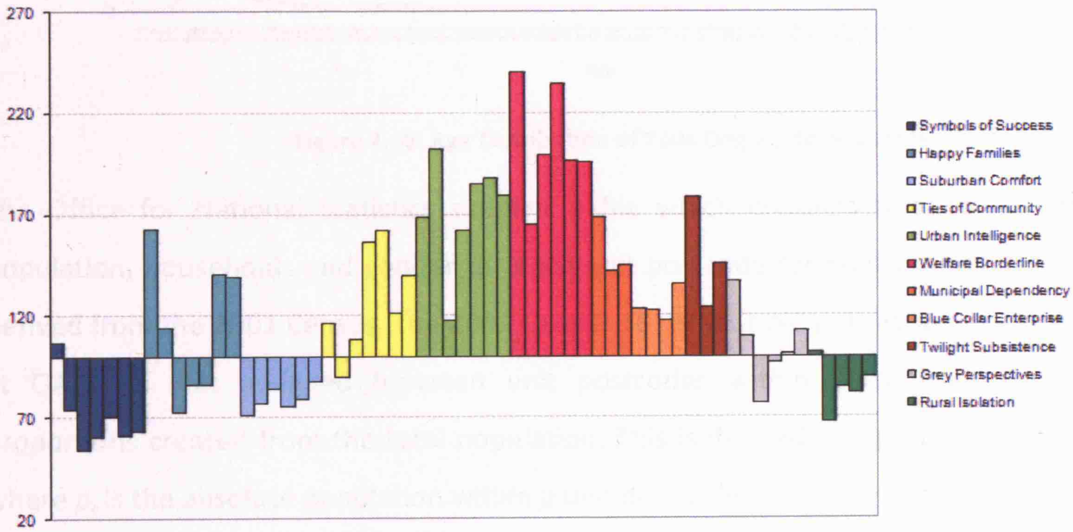


Figure 4.39: 2004 Degree Acceptance Profile Excluding 19 Years Old or Younger by Mosaic Neighbourhood Type (Source: 2004 UCAS Data)

Although Higher Education is not restricted to certain age cohorts, the majority of people who accepted places to study degrees in 2004 were aged between 17-25, with a peak at age 18 (see Figure 4.40). Thus the selection of base scores which are used to calculate neighbourhood participation rates affects the outcome significantly. Figure 4.16 showed the participation rate against the national population and by interpreting this statistics we define Higher Education as available for all; that is we should be equally worried about non participation across all age ranges. However, the majority of policy initiatives to extend participation to underrepresented groups focus on young participants, i.e. those students aged between 18-19 or under 21. In order to calculate a statistic which represents these policy target groups a base aligned with the age range 18-19 was derived. One issue with postcode level geodemographic indicators is that national coverage population data are not disseminated at this scale across different age ranges. Thus, a method of disaggregating Census data at OA level into unit postcodes was devised.

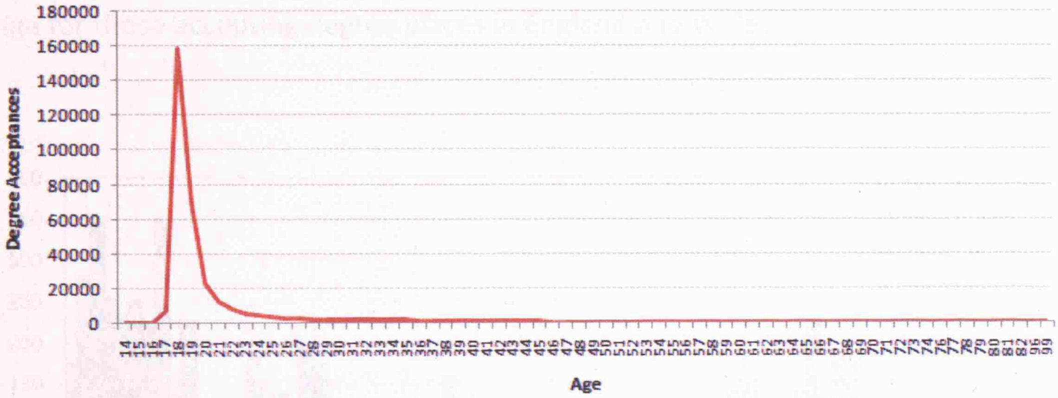


Figure 4.40: Age Distribution of 2004 Degree Acceptances (UCAS, 2004)

The Office for National Statistics supplies a file which contains frequencies for population, households and gender for each unit postcode for England and Wales derived from the 2001 Census. The 2001 Census 18-19 year old population available at OA level was assigned between unit postcodes within these areas using proportions created from the total population. This is derived using Equation (4.3) where p_i is the absolute population within a unit postcode, j the population of 18-19 year olds within Output Area y.

$$j_y \times \frac{p_{iy}}{\sum_{i=1}^{n} p_{iy}} \tag{4.3}$$

This algorithm was implemented in SAS[30] Macro language. The macro iterated through a list of all OA and using a mixture of SQL and SAS statements created a series of temporary tables in which calculations occurred. At the end of the application all temporary tables were amalgamated to create a new file of postcodes and their estimated frequencies of 18-19 year olds. These scores were deliberately not rounded, so a postcode could contain a proportion of a person. The postcodes were later aggregated into the 61 types by Mosaic, and as such unit postcode level rounding would have introduced an unnecessary margin of error.

In Figure 4.41 the estimated 18-19 year old base is compared with the same age range for those accepting degree places in England and Wales.

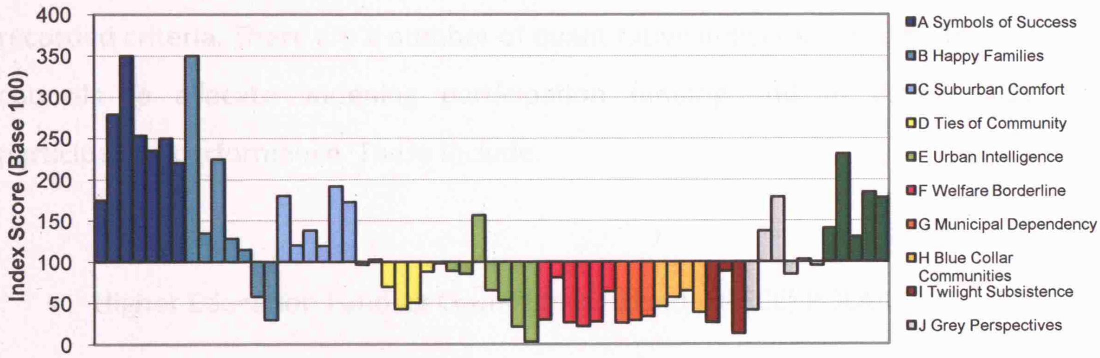


Figure 4.41: Participation Rates by Mosaic Group and Type in the 18-19 Age Cohort (Source: ONS & UCAS, 2004)

There is clear overrepresentation of Types falling within the Groups "Symbols of Success", "Happy Families", "Rural Isolation" and "Suburban Comfort" with the exceptions of Types "In Military Quarters" and "Burdened Optimists". The former of these Types contains very small counts and the index score may be anomalous. People living within "Burdened Optimists" areas are reasonably affluent, living in mortgaged property, and not having particularly high levels of education. The appearance of this Type in a typically high participation Group indicates a need for bespoke geodemographic classifications that better account for these within Group

[30] http//www.sas.com

differences. This theme is developed further in Chapter 6. A further Group demonstrating a heterogeneous Type distribution is "Grey Perspectives". The first Type within this group has a very small base and target score indicating that this index is probably unreliable. The second and third types within this group are called "Child Free Serenity" and "High Spending Elders". Both Types reside in affluent areas with high levels of educated residents. Students supplied from these neighbourhoods may therefore be the last in a line of siblings to attend university.

The concept of widening participation or extending access has already been introduced in Chapter 2 and the following analysis will evaluate this concept in quantitative terms through examination of current national indicators (see Chapter 3), and benchmarks for widening participation. Aggregate indicators for institutions provide a method of assessing performance on a range of directly and indirectly recorded criteria. There are a number of quantitative indicators used by the funding councils to allocate widening participation funding and to assess widening participation performance. These include:

- Higher Education Funding Council for England (HEFCE) POLAR
- National Statistics Socio-Economic Classification (NS-SEC)
- Proportion of state school students
- Geodemographics (Super Profiles)

Funding is given to institutions on the basis of their performance in widening access to young students (< 21 yrs) from disadvantaged backgrounds. The allocation of this money is currently calculated using a classification called POLAR. Five quintile bands were used to allocate £92.3m of funding in 2006/07 (HESA, 2006) and only those students in the lowest two participation bands qualify for an institution to receive funding. A key geographical problem with POLAR measurements of ward level participation rates that is not widely recognised is the heterogeneity which may occur within the ward. On average in England a ward contains 5500 people[31], and by assigning a single classification of participation rate assumes within ward homogeneity. Particularly in dense urban areas where the scale of socioeconomic differentiation often occurs at the level of the individual (Harris and Longley, 2005), these broad geographical aggregations fail adequately to measure micro level differentiation of population characteristics. This problem is illustrated in Figure 4.42 where the POLAR classification categorises the ward Gospel Oak into the 16-34% participation band. The postcode level Mosaic geodemographic classification is overlaid and shows the micro level heterogeneity within this area. Those unit postcodes in the north of the ward which border affluent Highgate are mainly categorised in Mosaic as "Symbols of Success", whereas those postcodes to the south of the ward are predominantly "Welfare Borderline". These geodemographic groups represent the highest and lowest participation neighbourhoods for Higher Education and illustrate how widening participation funding could potentially be incorrectly targeted through the use of large geographical aggregations in funding models. If these calculations were weighted with classifications using finer geographic scales then a more intelligent geographical model might be created.

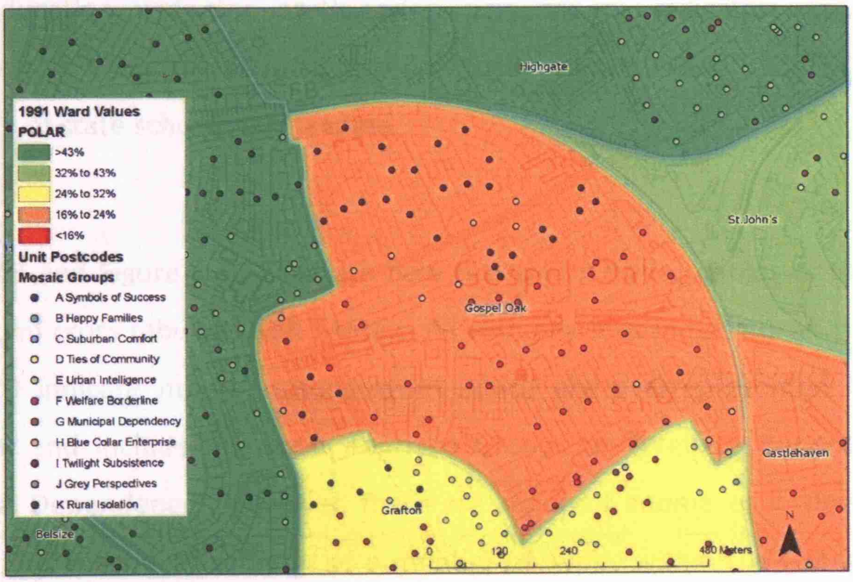


Figure 4.42: Mosaic and POLAR Classifications in the Areas Surrounding Gospel Oak in London (Source: Experian and HEFCE)

[31] http://www.statistics.gov.uk/geography/electoral_wards.asp

Although POLAR is used in the assignment of widening participation funding, it is not currently used in Higher Education performance indicators. Neighbourhood participation rates are currently measured using Super Profiles which is a geodemographic classification based on the 1991 Census of Population. These are outdated and have been superseded by more modern classifications such as the 2001 versions of Mosaic from Experian and ACORN from CACI.

Other national measures of widening participation performance used in the performance indicators include NS-SEC and the proportion of state school students, and although they may demonstrate spatial autocorrelation, they are attributed using individual characteristics and not geographical locations. As previously discussed in Section 3.3.2, the former of these classifications is based on the parental occupation of applicants under the age of 21, and when over the age of 21 their personal occupation. The proportion of state school students is derived using the UCAS classification of state and independent schools.

In order to better understand how neighbourhood classifications relate to both Higher Education widening participation performance indicators and funding mechanisms, a cross tabulation between Mosaic Types, POLAR, NS-SEC and the proportion of state schools was created.

Figure 4.43 and Figure 4.44 illustrate how the lowest two quintiles of the POLAR classification cross tabulate with Mosaic. As with previous index scores, those bars above 100 indicate neighbourhood types which are overrepresented in Higher Education, and include the most deprived Groups of "Welfare Borderline" and "Municipal Dependency". However, there are apparent anomalies in these results including the overrepresentation of a number of Types within the Group "Happy Families" – which based on their socio-economic characteristics, would have few restrictions on their propensity to participate in Higher Education. These results are created from all 2004 degree acceptances and as such illustrate a systematic problem within the POLAR classification.

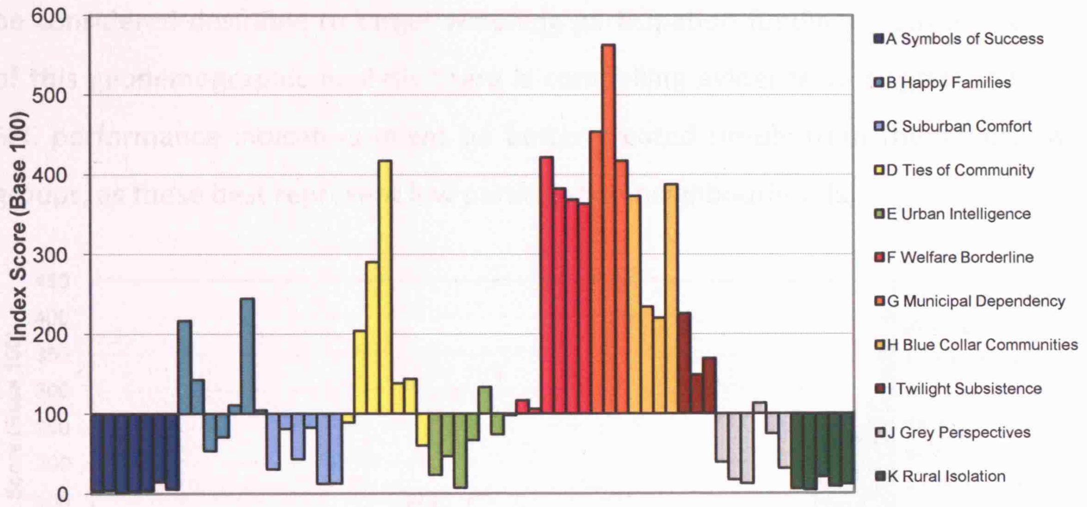


Figure 4.43: Very low (<16%) POLAR Participation Groups, by Mosaic Groups and Types (Source: HEFCE)

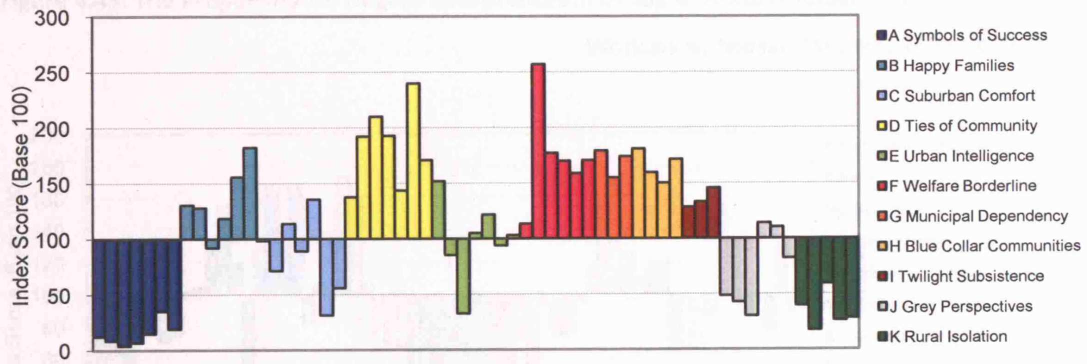


Figure 4.44: Low (16-24%) POLAR Participation Group by Mosaic Group and Type (Source: HEFCE)

The results from the cross tabulation for NS-SEC are shown in Figure 4.45, Figure 4.46, Figure 4.47 and Figure 4.48. Figure 4.45 shows an overrepresentation of the Group "Rural Isolation" and demonstrates how univariate classifications based on occupation interact with the geographic arrangement of typical industry types. These neighbourhood Types might be described as affluent, and as such should not necessarily be considered as areas where widening participation students might live. It could also indicate an inadequacy in Mosaic at capturing areas of rural poverty accurately. Similarly Figure 4.46 shows an over representation of mainly affluent neighbourhood Types, with the exception of "Blue Collar Communities". Figure 4.47 and Figure 4.48 show these students to have a higher propensity to live in less affluent neighbourhoods, and more in line with those areas where it might be considered desirable to target widening participation funding. From the results of this geodemographic analysis there is compelling evidence to suggest that NS-SEC performance indicators might be better created simply from the lowest two groups, as these best represent low participation neighbourhoods.

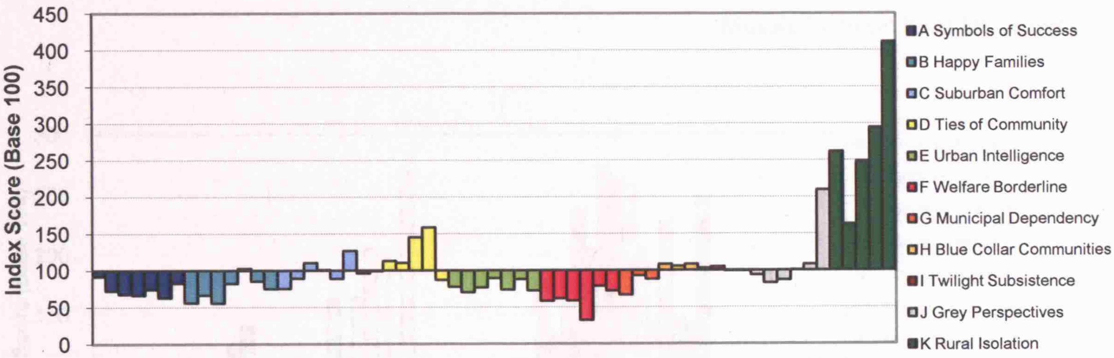


Figure 4.45: The Propensity for Degree Acceptances in Group 4: Small Employers and Own Account Workers by Mosaic (Source: 2004 UCAS Data)

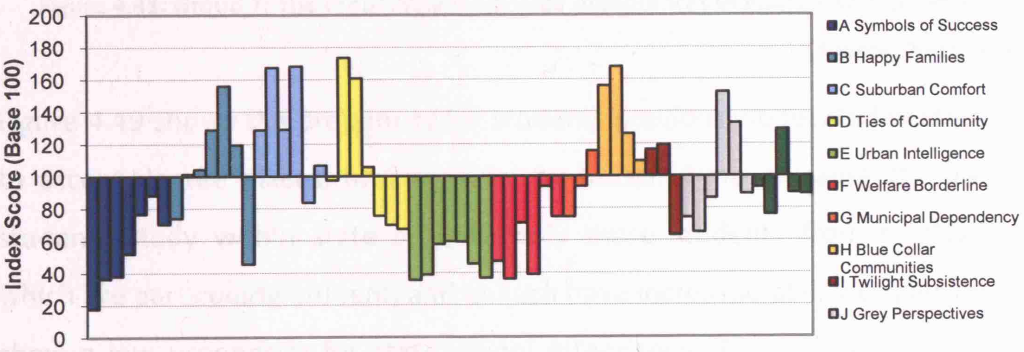


Figure 4.46: Group 5: Lower Supervisory and Technical Occupations (Source: 2004 UCAS Data)

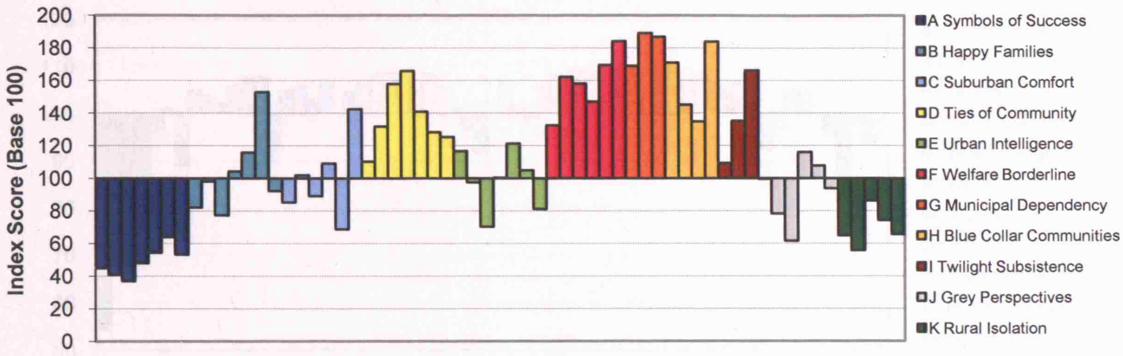


Figure 4.47: The Propensity for Degree Acceptances in Group 6: Semi-Routine Occupations by Mosaic (Source: 2004 UCAS Data)

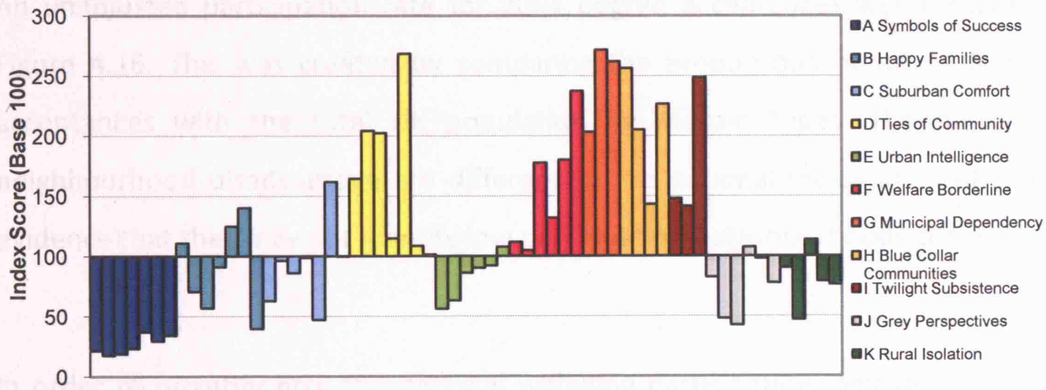


Figure 4.48: Group 7: The Propensity for Degree Acceptances in Routine Occupations by Mosaic (Source: 2004 UCAS Data)

Figure 4.49 shows the propensity for students previously educated in state schools to accept degree places, broken down by Mosaic Type. Because the majority of students study within state schools, only those students from neighbourhoods which are particularly affluent, and as such have increased ability to pay school fees show a low propensity for state school attendance. The differentiation between neighbourhood types for this particular indicator is therefore quite low.

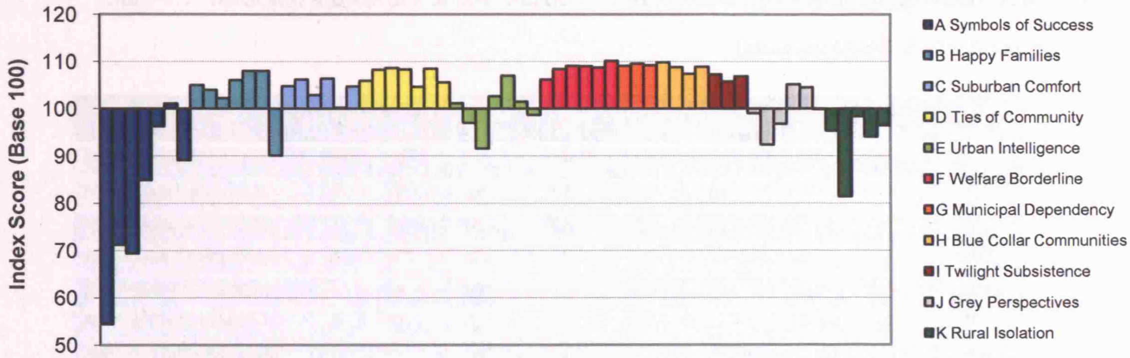


Figure 4.49: The Propensity for Degree Acceptances in State School Students by Mosaic (Source: 2004 UCAS Data)

An unadjusted participation rate for 2004 degree acceptances was presented in Figure 4.16. This was created by comparing the proportions of total UK degree acceptances with the total UK population by Mosaic Types. The patterns of neighbourhood disadvantage are different to the national indicators and provide evidence that they may not identify low participation neighbourhoods correctly.

In order to quantify how the national widening participation indicators discussed earlier (including <16%POLAR, 16-24% POLAR, NS-SEC 4, NS-SEC 5, NS-SEC 6 and NSSEC 7 and State Schools) relate to the actual participation rates at neighbourhood level (see Figure 4.41), the index ranking by the 61 Mosaic Types for each of these national measures was compared to the actual ranking of participation. Additionally, Experian provide a ranking of Mosaic Types by "Wealth" which is based on aggregated consumer income data. The rankings for these various measures are compiled in Table 4.7.

Table 4.7: National Indicators of Low Participation Ranked by Mosaic Neighbourhood Type (Source: UCAS, 2004 & Experian)

	HE	POLAR <16%	POLAR 16-24%	NSSEC 4.	NSSEC 5.	NSSEC 6.	NSSEC 7.	State School	Wealth
A01 Global Connections	15	59	58	28	62	59	60	61	47
A02 Cultural Leadership	3	57	59	49	59	61	62	59	56
A03 Corporate Chieftains	2	61	61	52	57	62	61	60	61
A04 Golden Empty Nesters	4	60	60	55	50	58	59	57	60
A05 Provincial Privilege	6	58	57	46	34	56	56	50	58
A06 High Technologists	5	47	49	56	31	53	58	41	59
A07 Semi-Rural Seclusion	9	55	55	37	42	57	57	56	57
B08 Just Moving In	1	16	26	60	40	44	25	30	26
B09 Fledgling Nurseries	21	21	28	53	23	37	45	35	48
B10 Upscale New Owners	8	38	39	61	22	47	48	38	55
B11 Families Making Good	23	35	30	40	11	33	36	25	44
B12 Middle Rung Families	26	26	17	19	7	27	21	16	43
B13 Burdened Optimists	42	12	6	36	15	13	19	17	25
B14 In Military Quarters	52	28	37	45	53	40	55	55	36
C15 Close to Retirement	12	43	43	44	13	43	47	31	54
C16 Conservative Values	24	32	32	29	4	34	34	23	50
C17 Small Time Business	19	40	40	11	10	41	40	36	46
C18 Sprawling Subtopia	25	31	24	20	2	30	32	22	45
C19 Original Suburbs	10	49	51	31	33	50	52	42	52
C20 Asian Enterprise	16	48	45	9	20	16	15	32	27
D21 Respectable Rows	30	30	23	25	24	28	31	26	35
D22 Affluent Blue Collar	27	17	5	24	1	21	14	15	34
D23 Industrial Grit	38	11	3	10	5	12	9	11	24
D24 Coronation Street	47	4	4	12	21	9	11	14	13
D25 Town Centre Refuge	45	22	22	8	36	17	22	33	18
D26 South Asian Industry	34	20	2	7	43	22	2	12	8
D27 Settled Minorities	29	37	12	34	44	23	26	28	21
E28 Counter Cultural Mix	32	44	19	42	61	25	30	40	20
E29 City Adventurers	35	39	41	51	55	38	49	47	39
E30 New Urban Colonists	17	53	50	43	49	49	46	54	49
E31 Caring Professionals	39	36	35	30	47	35	39	37	31
E32 Dinky Developments	44	23	29	47	48	24	37	20	30
E33 Town Gown Transition	59	34	38	32	52	32	35	39	23
E34 University Challenge	61	29	36	48	58	45	27	45	15
F35 Bedsit Beneficiaries	51	24	33	59	51	20	24	24	14
F36 Metro Multiculture	37	27	1	57	60	10	29	13	7
F37 Upper Floor Families	56	3	9	58	41	11	13	6	6
F38 Tower Block Living	58	7	14	62	56	14	20	7	1
F39 Dignified Dependency	54	9	16	41	26	6	12	10	5
F40 Sharing a Staircase	41	10	13	50	37	3	6	1	2
G41 Families on Benefits	57	2	8	54	38	7	10	5	3
G42 Low Horizons	53	1	18	27	25	1	1	3	4
G43 Ex-industrial Legacy	50	5	10	33	18	2	3	4	10
H44 Rustbelt Resilience	46	8	7	13	6	5	4	2	12
H45 Older Right to Buy	43	13	15	16	3	15	8	9	22
H46 White Van Culture	40	15	20	14	14	19	17	18	19
H47 New Town Materialism	49	6	11	18	19	4	7	8	11
I48 Old People in Flats	55	14	27	17	17	29	16	19	9
I49 Low Income Elderly	33	19	25	22	16	18	18	27	16

A Spearman's Rank Correlation (r_s) was calculated from these data to assess overall correlation between the actual 18-19 year old participation profile ("HE" Column in Table 4.7) and the national classifications where d_i are the differences between ranks x and y, and n are the frequency of types (see Equation (4.4)). Included in the table of ranks (Table 4.7) is a wealth rank. This ranking is provided by Experian in the supplementary material accompanying Mosaic, however in this instance the inverse ranking was used, and as such is a measure of low wealth and therefore should be positive correlated with indicators that show increased propensity in less affluent areas.

$$r_s = 1 - \frac{6 \times \sum d_i^2}{n \times (n^2 - 1)}$$

(4.4)

Table 4.8 shows the results of Spearman's Rank Correlation.

Table 4.8: Spearman's Rank Results (Low Actual Participation)

Classification	Spearman's Rank
POLAR <16%	0.77
POLAR 16-24%	0.66
NS-SEC 4.	0.01
NS-SEC 5.	-0.03
NS-SEC 6.	0.77
NS-SEC 7.	0.69
State School	0.64
Wealth	0.83

Of these indicators the highest correlation with actual participation rates are tied between POLAR <16% and NS-SEC 6 with a score of 0.77. One would expect a high correlation with the lowest POLAR quintile as it is the only national indicator that directly measures low participation. It is also reassuring to find quite a strong correlation between the POLAR 16-24% class and neighbourhood level participation. Oddly NS-SEC 7, which is less "affluent" than NS-SEC 6, has a lower correlation with actual neighbourhood participation rates, indicating an issue with this particular indicator. As might be expected given the index profiles presented in

Figure 4.45 and Figure 4.46, NS-SEC 4 and 5 shows very low correlation with participation rates, with correlations of 0.01 and -0.03 respectively. State school attendance shows quite a strong correlation of 0.64, although this is one of the lower correlations. Neighbourhood wealth has a correlation of 0.83 which is higher than any of the national benchmarks, indicating a very strong relationship between financial means and participation in Higher Education. The interesting conclusion from this analysis is that the indicators used in Higher Education as benchmarks do not correlate highly with neighbourhood participation rates, providing evidence that these measures may be equally ineffective in performance monitoring and funding assignment. Furthermore, the best performing indicator is one which is not based on public sector data sources, suggesting the need to access indicators of wealth, perhaps through public/private sector data sharing partnerships.

4.5 Conclusion

This chapter has presented an exploratory data analysis predominantly using index scores and has shown how access to Higher Education is both spatially and socially heterogeneous. The distance that accepted applicants travel to start their degrees has been shown to be stratified spatially by both course and institution. The few locations that a course is offered, the further (on average) students travel to accept these places as they are offered at a restricted selection of potential institutions. Institutions with high entry qualifications were shown to attract students from a broader area. When average tariff scores were examined by neighbourhood types, higher attainment was recorded in students from more affluent neighbourhoods, thus having fewer restrictions on travel, both socially and economically. A model was created which estimated the 18-19 year old base population in England and Wales at unit postcode level and was compared to the same age range accepting Higher Education places in 2004. This was aggregated to create a measure of neighbourhood participation. This ranking of neighbourhood level young participation was compared to those rankings created by the classification used by HEFCE and HESA to allocate widening participation funding and assess benchmarked performance. Variable performance was found, with high correlation with the lowest POLAR and NS-SEC 6. Some issues requiring further investigation were the lower correlation with NS-SEC 7 & 4. The highest correlation was recorded using the Experian "wealth" ranking.

5

DECISION SUPPORT AND PROFILING FOR HIGHER EDUCATION

5.1 Data Complexity and Decision Making

Chapter 2 concluded that the educational data economy is complex and provided by multiple organisations across a variety of sectors. The availability of relevant data to Higher Education Institutions for decision making is critical if they are to utilise public funding most effectively. Currently Higher Education institutions do not have a centralised, coordinated and freely available source of such intelligence, but can derive it from a range of commercial and non commercial data providers. Following a review of the disparate services that are currently available to institutions wishing to develop an information base in order to extend access, this chapter introduces a pilot educational decision support service which aims to centralise the dissemination of useful information in a structured and digestible format. The application code is presented in the Appendix Figure 12.11. A case study using UCL data is embedded within this presentation to demonstrate how the analysis created is of practical relevance. The main services currently available to institutions are reviewed in Sections 5.1.1 - 5.1.5.

5.1.1 UCAS Analytical Services

The largest and main centralised provider of analysis and information to the Higher Education sector is the Universities and Colleges Admissions Service (UCAS), that has developed a series of products and services which aim to meet institutional information needs. However, in reality these fall somewhat short of the kinds of analysis which are required in an increasingly market orientated sector. Current products and services include:

- UCAS online statistics
- Annual datasets
- Application tracker
- Bespoke analysis query service

UCAS statistics is an online query tool which serves cross tabulations of some pre-generated statistics for large aggregate groups, e.g. frequency of ethnic minority groups attending courses of Higher Education. This service is useful, however it does not provide neighbourhood classification, nor does it allow aggregation by any spatial unit. Annual datasets are sent to each institution which provides a summary of an institution's application cycle as a range of HTML summary tables and Excel pivot tables. These data provide some useful information to institutions such as the other institutions to which their accepted applicants applied using their other UCAS choices, i.e. competitor institutions. However, these data are not in an easily manipulated format, nor are they particularly easy to interpret without knowledge of pivot tables. Application tracker is the latest service provided to institutions and requires an annual subscription. This service provides institutions with live summary information of applications to their institution and basic competitor analysis. The service is useful, however missed the opportunity to flag applications that could be considered to meet widening participation targets, thus orientating it purely as a marketing and operational support tool. The final service provided by UCAS and charged at a cost recovery rate is the bespoke analytical query service. Users can request an individual query based on any of those data collected from UCAS and is generated manually. The overall advertising of all these services is poor and the output (perhaps with the exception of the bespoke query service) does not meet the analysis requirements of institutions seeking to effectively target widening participation initiatives or to market courses in Higher Education.

5.1.2 Higher Education Funding Council for England (HEFCE)

The Participation of Local Areas (POLAR) classification as discussed in Section 4.4.3 is provided by HEFCE through a series of PDF[32] files which can be downloaded and browsed. This is probably the best attempt at integrating areal data from multiple sectors into a map which is publicly available to download from a centralised source. POLAR 1998 ward level data are provided and overlaid with school GCSE attainment data (See Figure 5.1).

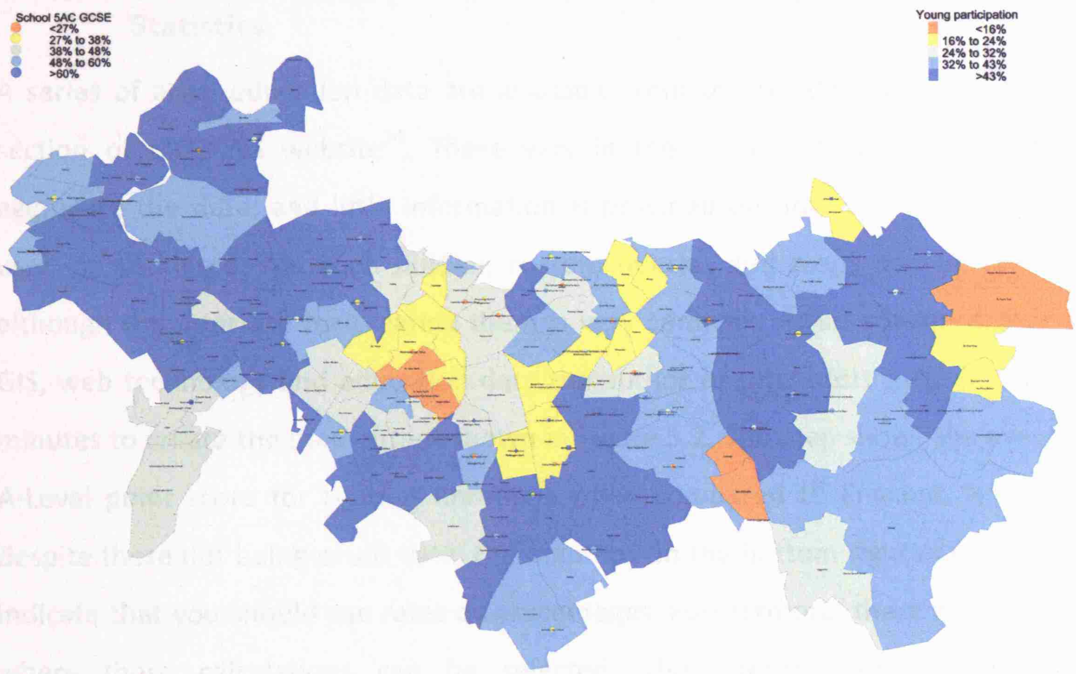


Figure 5.1: A POLAR Map Provided by HEFCE for South London

5.1.3 Department for Children, Schools and Families (DCSF)

The DCSF has recently redesigned their school performance website using an improved layout and featuring Google map integration. These data could be used by an institution to manually check the attainment of a pupil against their school and local authority averages.

[32] http://www.adobe.com/uk/products/acrobat/

5.1.4 LG01[33]

LG01 are the most successful Higher Education targeting software provider in the UK with over 19 institutions currently subscribed to their services. These institutions provide LG01 with a series of operational datasets which are housed on the LG01 secure server and may be queried using software tools. The LG01 software suite includes LINK which is an analysis and geographic profiling tool, PROFILE which appends a series of widening participation labels to applicant data and REACH for monitoring outcomes of outreach activity. LG01 is unfortunately a monopoly provider and as such universities pay a premium price for analysis of data which they already own.

5.1.5 Office for National Statistics (ONS) and Neighbourhood Statistics

A series of areal education data are available from the Neighbourhood Statistics section of the ONS website[34]. These vary in the areal units that are used to aggregate the data, and little information is provided on how the data might be used or visualised. Although hidden, custom queries and maps can be created, although the interface for creating them is very complex. As an advanced user of GIS, web technology and education data the author of this thesis took around 30 minutes to create the map output shown in Figure 5.2. The map shows the average A-Level point score for London Boroughs when compared to England. However, despite these not being count data, the map tips on the bottom right of the screen indicate that you should use rates or percentages. Furthermore, there is no option where these calculations can be selected. The intervals are automatically generated, and you have the option of specifying three, four or five class boundaries. The services provided by the ONS have the potential to be very useful; however severe usability issues limit their impact.

[33] http://www.lg01.com/
[34] http://neighbourhood.statistics.gov.uk

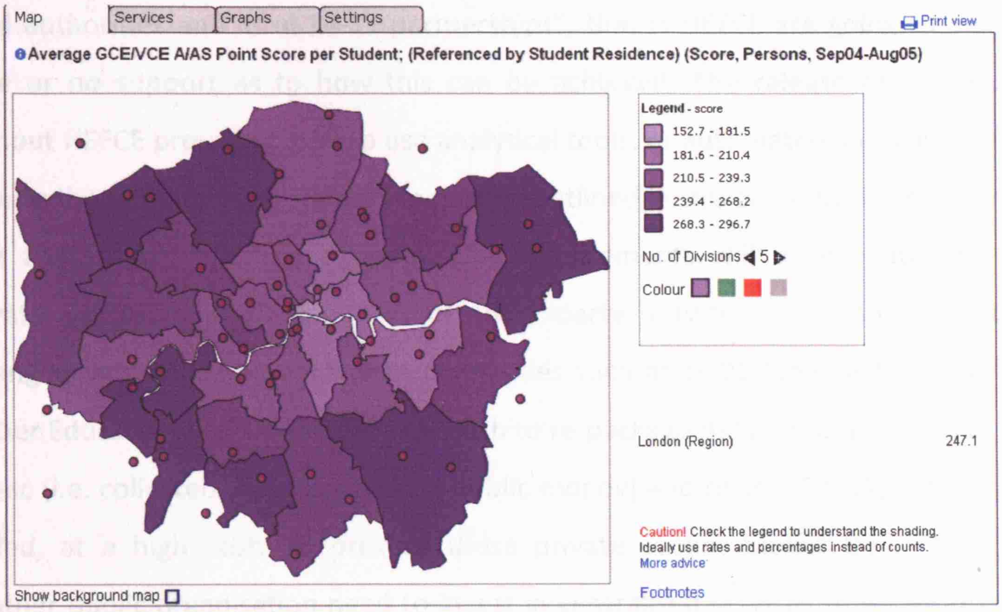


Figure 5.2: Neighbourhood Statistics Map Output

5.1.6 Data Services Strategy

The 2007 HEFCE publication titled "Higher Education Outreach: Targeting Disadvantaged Learners" discussed strategies where by widening participation groups could be targeted. The report concluded:

> *"From autumn 2007, HEFCE will make available data sets for small areas grouped by the rates of young participation in HE, and for small areas grouped by relative deprivation (drawing on the Index of Multiple Deprivation for super output areas). These data will provide all Aimhigher partnerships and HE providers with important contextual information at area level. Details for learners involved in WP activities can also be compared with the HEFCE data to check the accuracy of the targeting process. Targeting at area level should draw on the expertise of local authorities and local 14-19 partnerships, where appropriate localised professional advice or additional data can be obtained. Fischer Family Trust data, and data on free school meals and educational maintenance allowances, are also useful. We will not require Aimhigher partnerships or HE providers to adopt a single model, but we would expect them to establish a well-informed, systematic and transparent method of combining data on deprivation, attainment and participation in order to satisfy the demands of the first stage in targeting."*

HEFCE (2007b:26)

This statement highlights that HEFCE is going to provide some local area data which can be used by Aim Higher Partnerships[35] and institutions to target resources, however these analyses "targeting at area level should draw on the expertise of

[35] http://www.aimhigher.ac.uk/

local authorities and local 14-19 partnerships", that is HEFCE are going to provide little or no support as to how this can be achieved. The release of these data without HEFCE providing easy to use analytical tools, or automated analysis services such as the pilot decision support tool to be outlined in this chapter, there is danger that these additional and essential user requirements will be exploited by the private sector through the sale of third party services. It is these sort of arrangements which have allowed companies such as LG01 (See Section 5.1.4) in Higher Education and Dr Foster[36] in Health to re-package data already owned by the public (i.e. collected and stored using public money) and re-sell it back, albeit value added, at a high cost. To prevent these private sector exploitations HEFCE or another public organisation need to invest in centralised services to perform these analysis tasks for institutions on a cost recovery basis. This would be a far more cost efficient solution than each institution investing in their own analysis from a range of different sources.

5.2 Introducing and Initialising an Educational Market Profiler

Section 5.1 has outlined how those data currently used for decision making in Higher Education are available from multiple sources and of varying degrees of usability. The Education Market Profiler (EMP) application presented in the remaining part of this chapter provides a solution that UCAS could deploy to integrate both school and Higher Education data into those analysis which enable institutions to strategically market courses and extend access to underrepresented groups. This is a unique contribution to the educational data economy as it integrates a series of sector data for the first time and also allows institutions to benchmark their performance against competitor institutions across a range of variables.

Following discussion with the UCAS analytical services team and through examination of those common queries which are raised through direct stakeholder

[36] http://www.drfoster.co.uk/

contact, a list of user requirements was created. These consisted as a series of questions that a profiling application should address:

- Where is my core recruitment region?
- How can I target mailing lists to appropriate students?
- Which schools do I need to engage with for widening participation?
- How do I assess my participation profile?
- How can I benchmark department and faculty performance at widening participation within my institution?

UCAS Analytical Services department are users of the SAS[37] statistical software and as such EMP was written as a series of SAS modules which could integrate with the UCAS annual datasets created in the SAS database format. The application was designed so that by adapting a series of input parameters it could generate analysis for any UCAS institution. The modules which make up EMP included:

- Data Preparation Module
- Geographic Analysis
    - Distance Analysis Module
    - Postcode Geography Module
    - Postcode Extraction Module
    - Regional Analysis Module
- Schools Analysis
    - PLASC Schools Module

[37] http://www.sas.com/

    - UCAS Schools Module
    - Stitching Module

- Institutional Benchmarking
    - Institutional Profile Module

5.2.1 Setting up the Educational Market Profiler

Each EMP module will be discussed throughout the rest of this chapter and flow diagrams will be used in order to explain how the SAS code creates the output files. A case study for University College London is presented to demonstrate the output from this decision making tool. Before the application was run it was necessary to define a series of variables which are shown in Figure 5.3.

```
*/Library Names & Initial Variables;
Libname Look 'Y:\Alex\MOSAIC';
Libname _02 'R:\2002';*/ 2002 Data;
Libname _03 'R:\2003';*/ 2003 Data;
Libname _04 'Y:\2004';*/ 2004 Data;
Libname Distaste 'R:\Distance_data';
Libname workdir 'Y:\Alex\SAS Working Data';
Libname poldir 'Y:\Alex\POLAR Data\POLAR SAS';
Libname Uschools 'Y:\Alex\Schools';*/ UCAS Schools Data;
Libname Pschools 'Y:\Alex\Schools\PLASC';*/ Dfes Schools Data (PLASC);
Libname schlook 'Y:\School File Update';*/ school Dfes - UCAS lookup file;
%LET MosCSV="Y:\Alex\Mosaic Files\UCAS.csv";*/Experian Lookup File;
%LET AFPD="Y:\Alex\Lookup\Postcode Data Files\pcluts_2004nov\Data\afg04nov.csv"; *Location of
AFPD File Directory;
%LET GISOUT="Y:\Alex\Geodem_app\GIS Output"; *GIS File Output Directory;
%LET STATOUT=Y:\Alex\Geodem_app\Stat Output; *Stat File Output Directory;
%LET PostSTAT=Y:\Alex\Geodem_app\Stat Output\Postcodes; *Postcodes Output Directory;
%LET CompCSV="Y:\Alex\Geodem_app\input files\comp\Compet.csv";*/Competitor CSV File;
%LET AgregCSV="Y:\Alex\Geodem_app\input files\agreg\agreg.csv";*/Aggregation CSV File (Crse >
Faculty / Department);
%LET inst_02=U80; *Institution Code 02;
%LET inst_03=U80; *Institution Code 03;
%LET inst_04=U80; *Institution Code 04;
```

Figure 5.3: SAS Code – Set up Initial Variables

The SAS libraries are used to locate SAS datasets, in a very similar way to file directories in the Windows operating system. These libraries included the location of the UCAS main datasets for 2002 – 2004, location of spatial co-ordinates for UCAS Higher Education institutions, a working directory where outputs would be stored, a POLAR directory storing the POLAR classification lookup tables, a library of UCAS schools data, a library of DCSF PLASC data and a library containing the lookup file used to amalgamate UCAS and DCSF data. In addition to the libraries, variables were defined for the location of the Mosaic geodemographic classification, the All

Fields Postcode Directory (AFPD), a GIS file output directory, a statistics file output directory, a postcode analysis output directory, the location of a CSV file containing competitor details, a file which links institution course codes in departmental and faculty aggregations and finally the institution codes for 2002 – 2004. A separate code was defined for each year as in rare occurrences an institution may change their establishment code between years as discussed in Section 2.3.

Once the above statements were run and variables assigned, it is possible to highlight the code in the SAS application window for a chosen module and to run each individually. However, certain modules have dependencies on the outputs of other modules. The input requirements for each module are detailed on the flow charts presented throughout the following section. The outputs from the application were all CSV files which were imported and formatted using pre-prepared Excel workbook with visual basic code. These files were copied into the directory specified by the "STATOUT" variable.

5.2.2 The Data Preparation Module

The Data Preparation module (See Figure 5.4) takes the 2002 – 2004 UCAS datasets and attaches a series of variables that are required throughout the analysis. In order to attach these variables several external CSV files are required, including:

- Experian Mosaic Classification Directory:
    - Postcode, Easting, Northing, Mosaic Type, Mosaic Group.
- All Fields Postcode Directory:
    - Postcode, Country, 1991 ward, 1998 ward, Output Area, Lower Level Super Output Area, CAS ward.

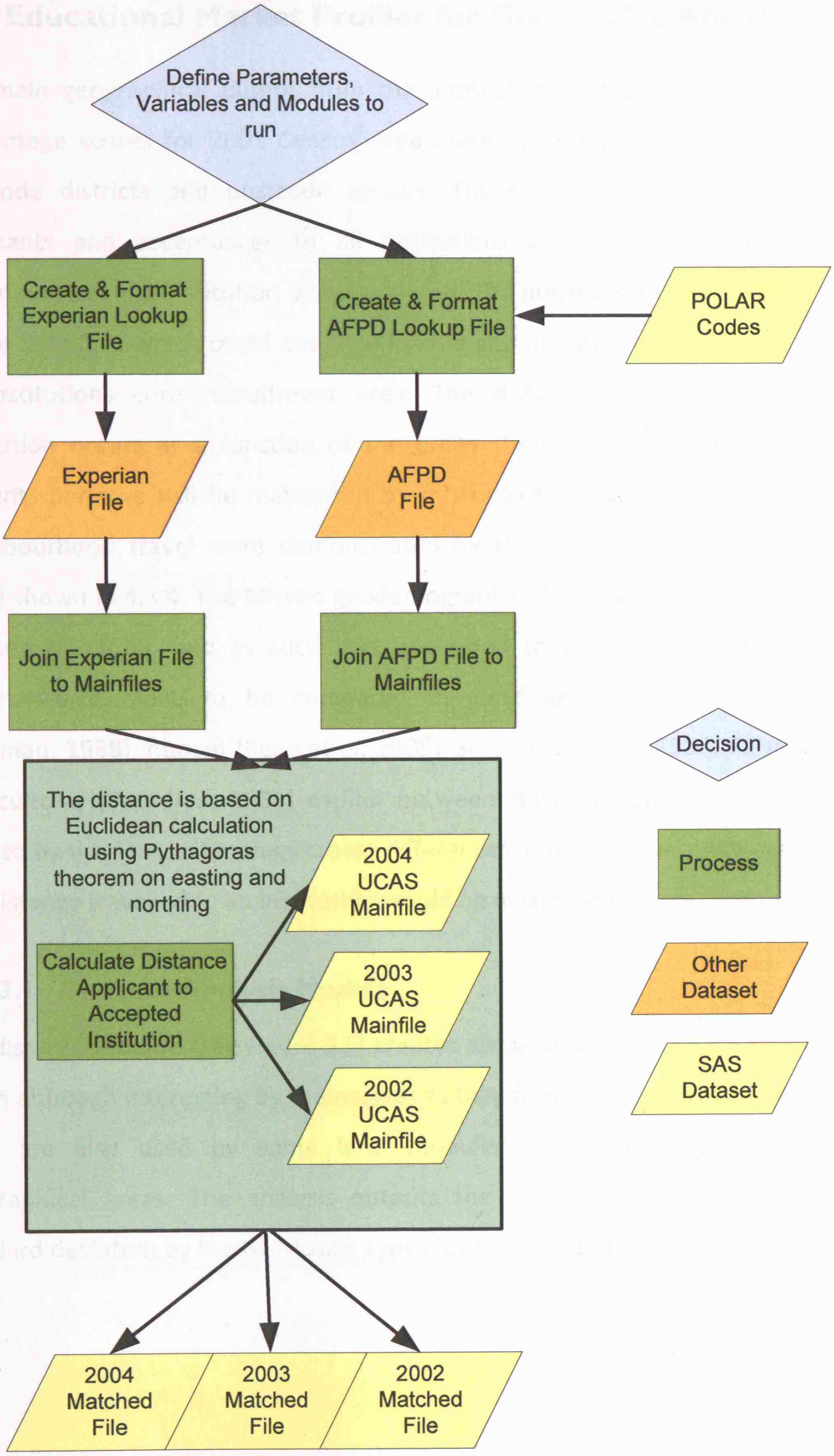


Figure 5.4: Data Preparation Module

5.3 Educational Market Profiler for Geographic Analysis

The main geographical output from the application was a series of counts and percentage scores for 2001 Census Area Statistics (CAS) wards, postcode areas, postcode districts and postcode sectors. These scores consisted of the total applicants and acceptances to all institutions and the total applicants and acceptances to the institution being analysed. The purpose of these outputs was to create data files which could easily be loaded and mapped in GIS tools to visualise an institutions core recruitment area. The distance applicants travel to an institution occurs as a function of the utility (Hensher and Johnson, 1981) the students perceive will be maximised by this choice. These aggregate patterns of neighbourhood travel were demonstrated by the distance decay charts (Lösch, 1954) shown in 4.3.4. The Mosaic geodemographic classification from Experian was licensed by UCAS, and as such was appended to the applicant data enabling acceptance distance to be compared by neighbourhood. Variable economic (Coleman, 1988), human (Reay *et al*, 2005), social (Coleman, 1988; Putnam, 1995) and cultural (Bourdieu, 1986) capital between those people living within areas defined by the Mosaic typology create differences between how utility in relation to the distance travelled to an institution would be maximised.

5.3.1 Distance Analysis Module

The distance module (See Figure 5.5) creates distance scores for each Mosaic Type, which although interesting by themselves as they form the geodemographic market area, are also used by some later modules to provide target analysis for geographical areas. The analysis outputs the minimum, maximum, mean and standard deviation by the 61 Mosaic Types for the selected institution.

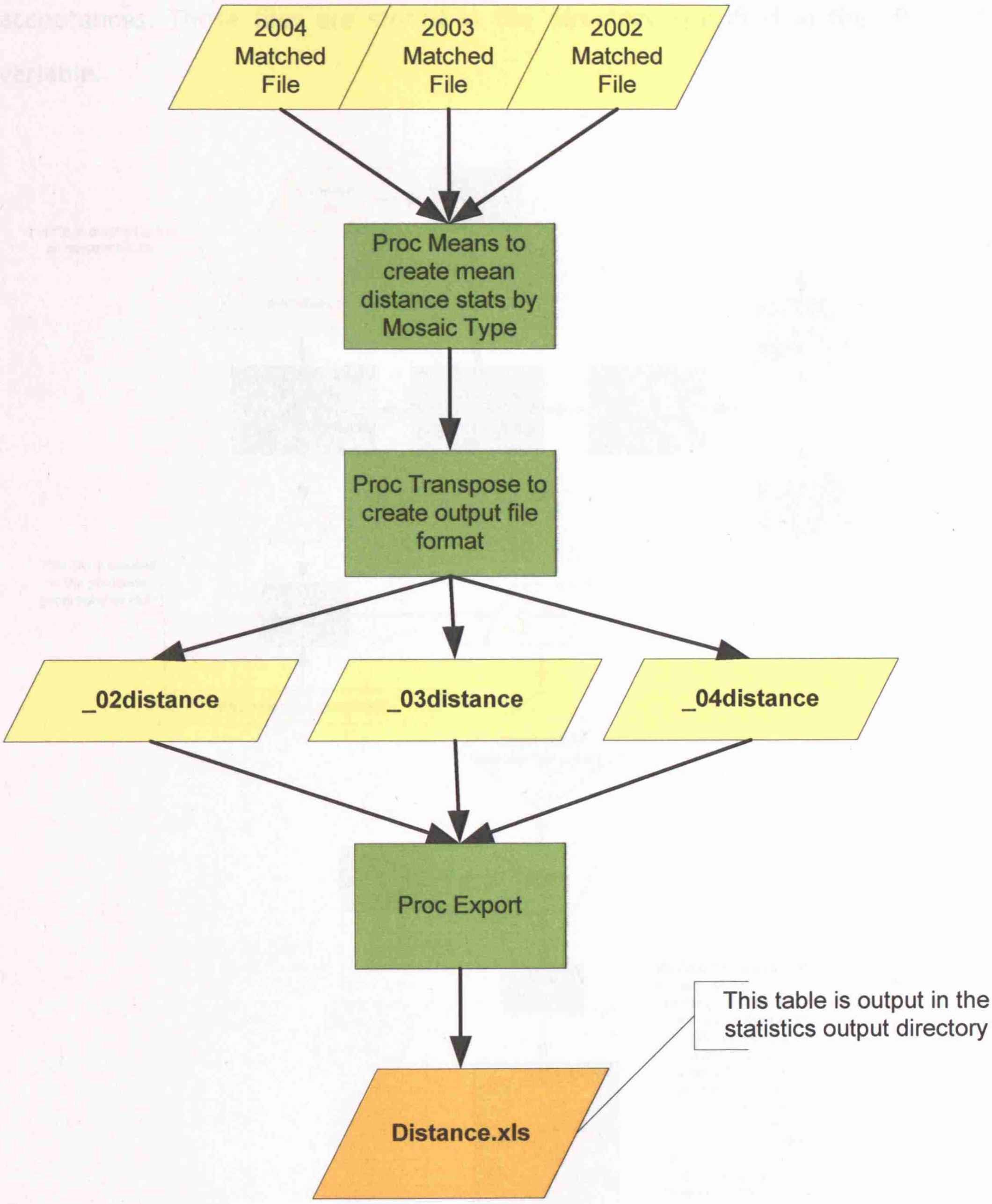


Figure 5.5: Distance Module

5.3.2 Postcode Extraction Module

This module (See Figure 5.6) creates CSV extracts based on the average distance buffer scores created in the Distance Analysis Module (See Section 5.3). All applicants are compared to these buffers and a separate CSV file is created for the 61 distance buffers for each Mosaic Type. Each file contains those total postcodes, total applicants and acceptances and selected institution applicants and acceptances. These files are stored in the directory specified in the "PostSTAT" variable.

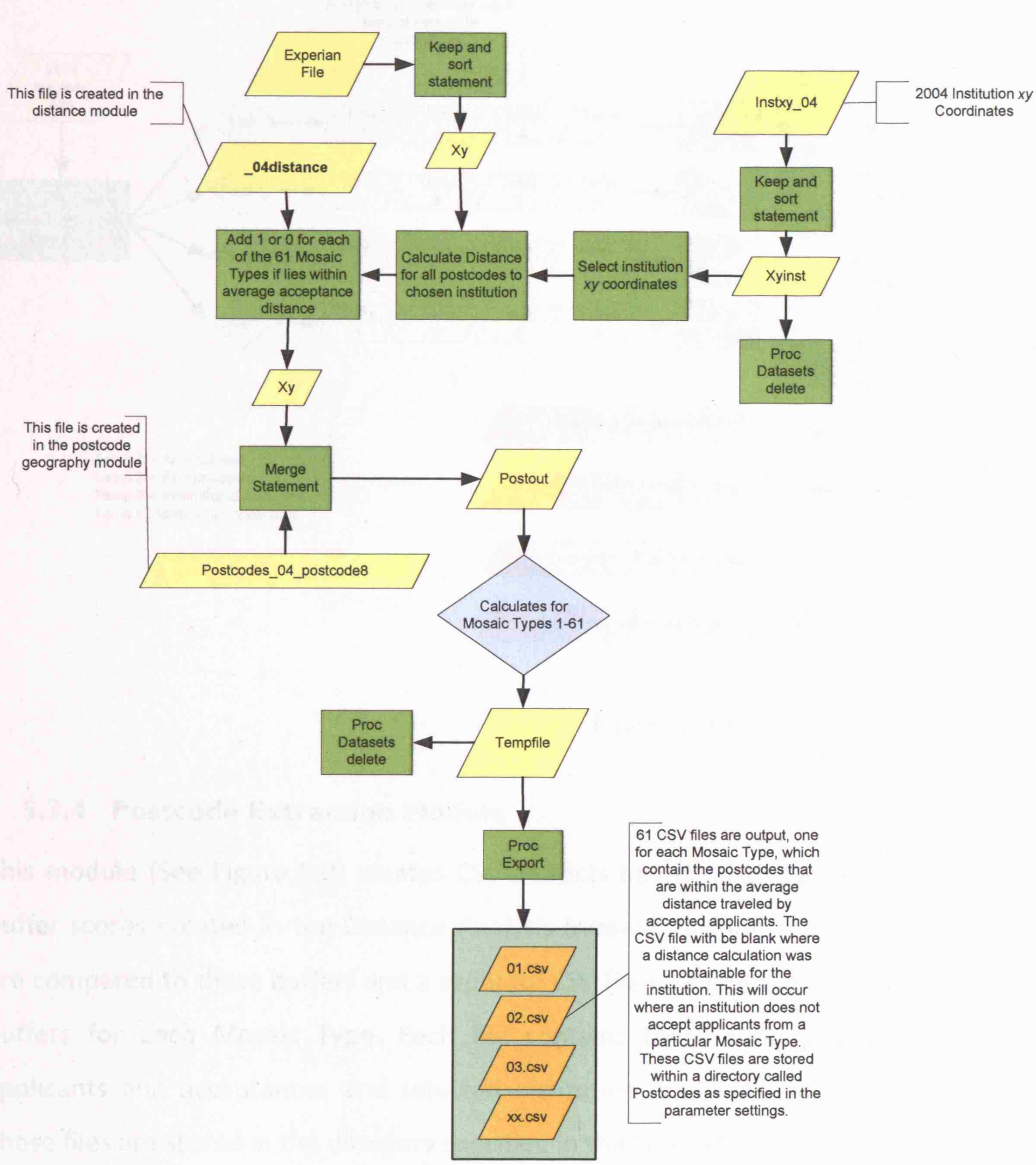


Figure 5.6: Postcode Extraction Module

5.3.3 Postcode Geography Analysis Module

This module (See Figure 5.7) counted the number of acceptances and applicants within each CAS Ward. These counts were created for both the total, and for the selected institution population using the 2004 data. These analyses are created for

Postcode Areas, Districts, Sectors and Units. The Excel output files are suitable for post processing in a GIS and linked with boundary data.

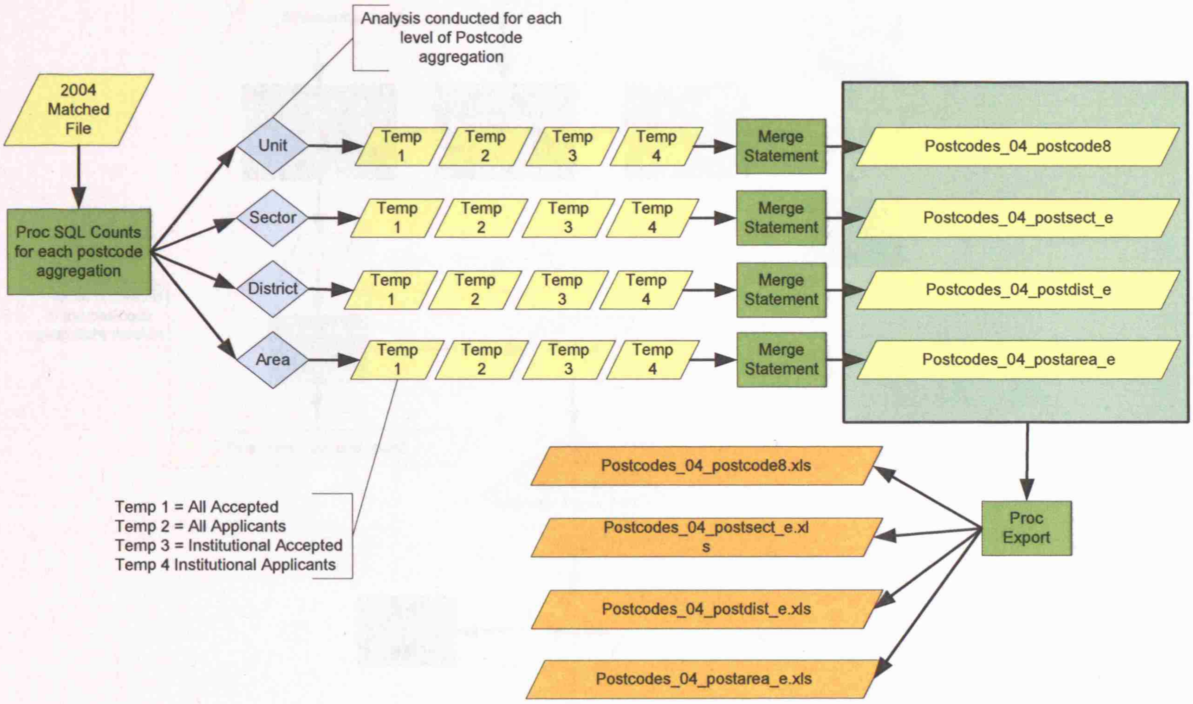


Figure 5.7: Postcode Geography Module

5.3.4 Postcode Extraction Module

This module (See Figure 5.8) creates CSV extracts based on the average distance buffer scores created in the Distance Analysis Module (Section 5.3). All applicants are compared to these buffers and a separate CSV file is created for the 61 distance buffers for each Mosaic Type. Each file contains those total postcodes, total applicants and acceptances and selected institution applicants and acceptances. These files are stored in the directory specified in the “PostSTAT” variable.

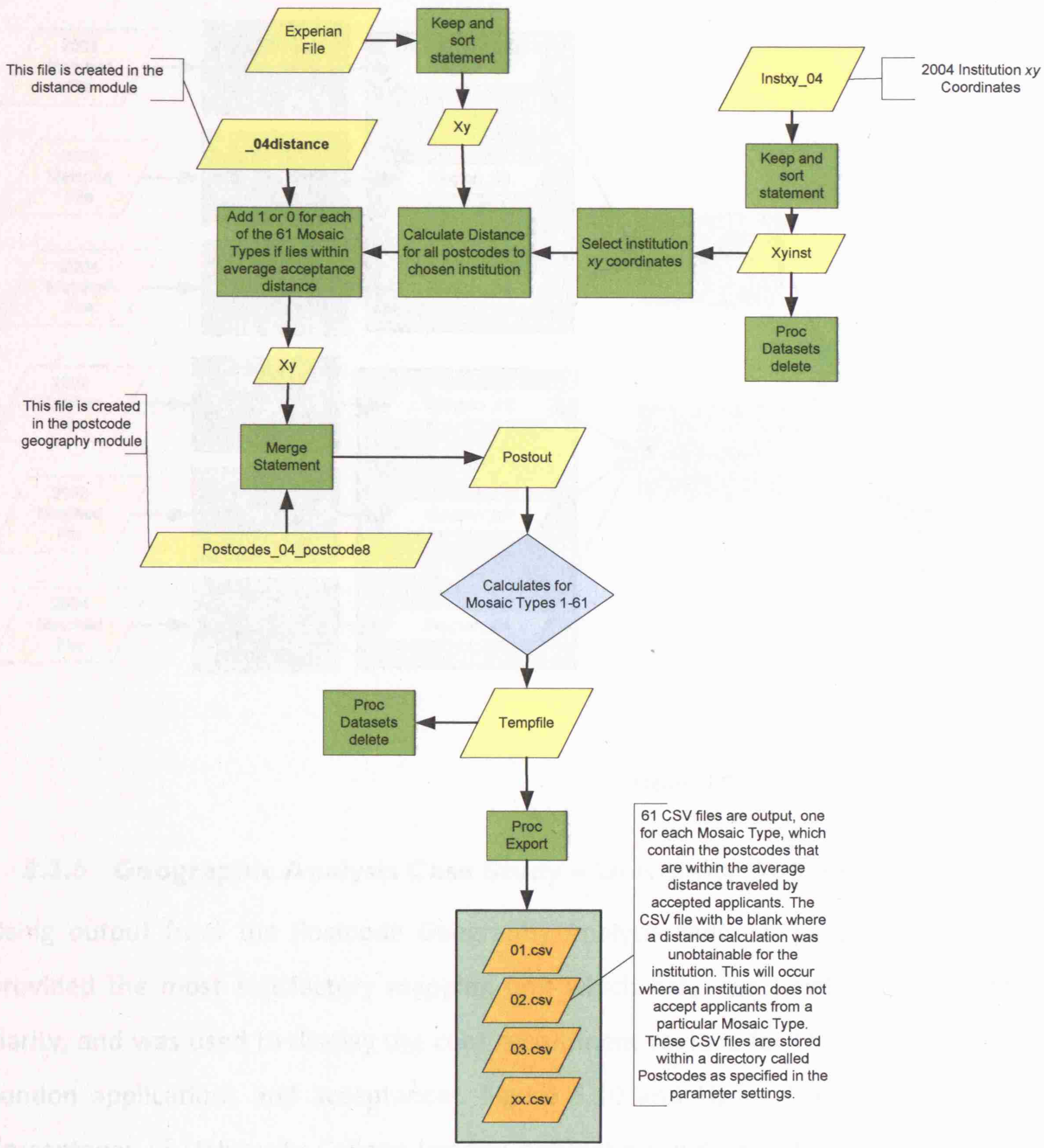


Figure 5.8: Postcode Extraction Module

5.3.5 Regional Analysis Module

The regional analysis module (See Figure 5.9) counts and sums the number of acceptances and applicants between 2002 and 2004. Using multiple years of data it creates a more robust estimate of an institutions market area. The Excel output from this analysis can be manipulated and then input into ArcGIS.

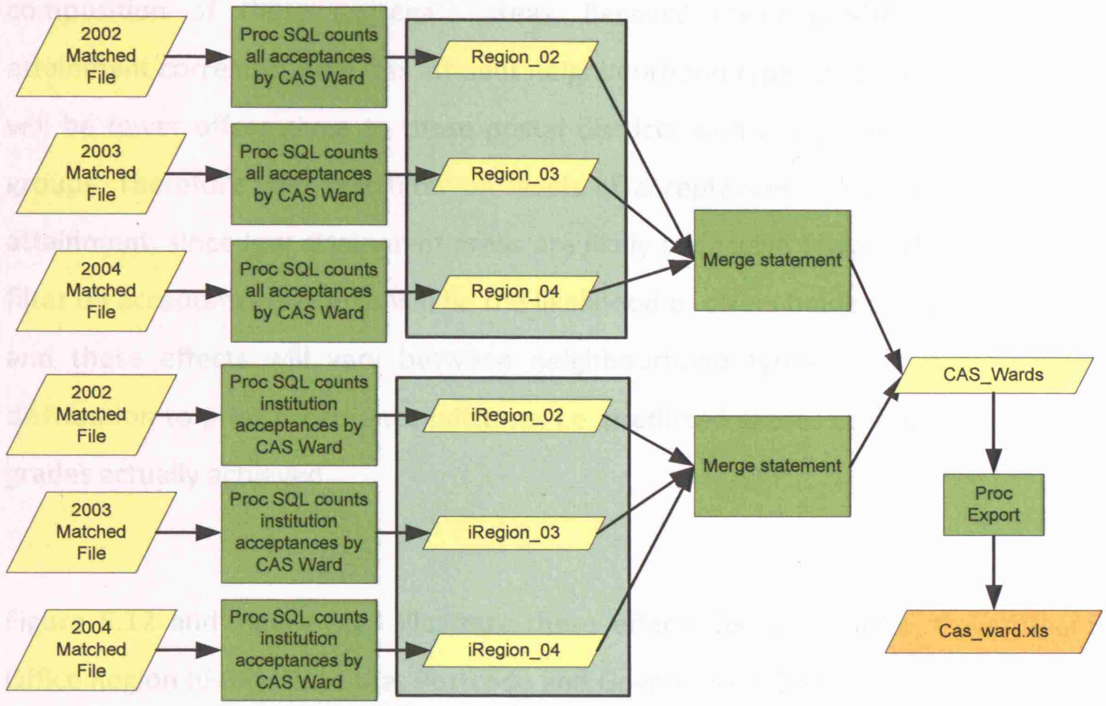


Figure 5.9: Regional Analysis Module

5.3.6 Geographic Analysis Case Study – University College London

Using output from the Postcode Geography Analysis Module, postcode districts provided the most satisfactory mapping unit which balanced level of detail with clarity, and was used to display the core recruitment regions for University College London applications and acceptances. Figure 5.10 and Figure 5.11 compare the percentages of University College London application and acceptance frequencies against a base of all applicants and acceptances. Figure 5.10 shows that University College London has a core applicant market centred on London and across the South East of England. There is a steady decline in the share of applications beyond this region, and particularly across Scotland. Figure 5.11 shows how University College London acceptances follow a similar regional pattern. This spatial distribution can be attributed to a combination of where offers are made regionally, given that this is not even across the UK for University College London (See Figure 5.10), and also where people are more likely to accept offers. Thus, the propensity to accept places from within any given Postal Districts will be dependent on both the proximity to University College London, and also the neighbourhood composition of these aggregate areas. Because lower pre-HE achievement/ attainment correlates with less affluent neighbourhood types (see Chapter 4), there will be fewer offers given to those postal districts with a high incidence of these groups. Therefore, the first filter on levels of acceptances will be levels of prior attainment, since low attainment areas are likely to receive fewer offers. A second filter on acceptance patterns will be the likelihood of offers being met by applicants and these effects will vary between neighbourhood types and have a similar distribution to prior acceptance patterns, i.e. predicted grades correlate with those grades actually achieved

Figure 5.12 and Figure 5.13 illustrate these effects for the London Government Office Region (GOR). Note that Postcode and Government Office Region boundaries are not coterminous. Furthermore, the Postal Boundary data are reasonably old, and as such some areas have changed their boundaries. Both maps have a light area to the North West of London[38]. These are Postal Districts WD1 and WD2 which no longer exist, having been renamed in June 2000.

[38] See ftp://ftp.royalmail.com/Downloads/public/ctf/rm/PU_Historic_Data_Web.pdf for explanation of possible inaccuracies in these boundaries.

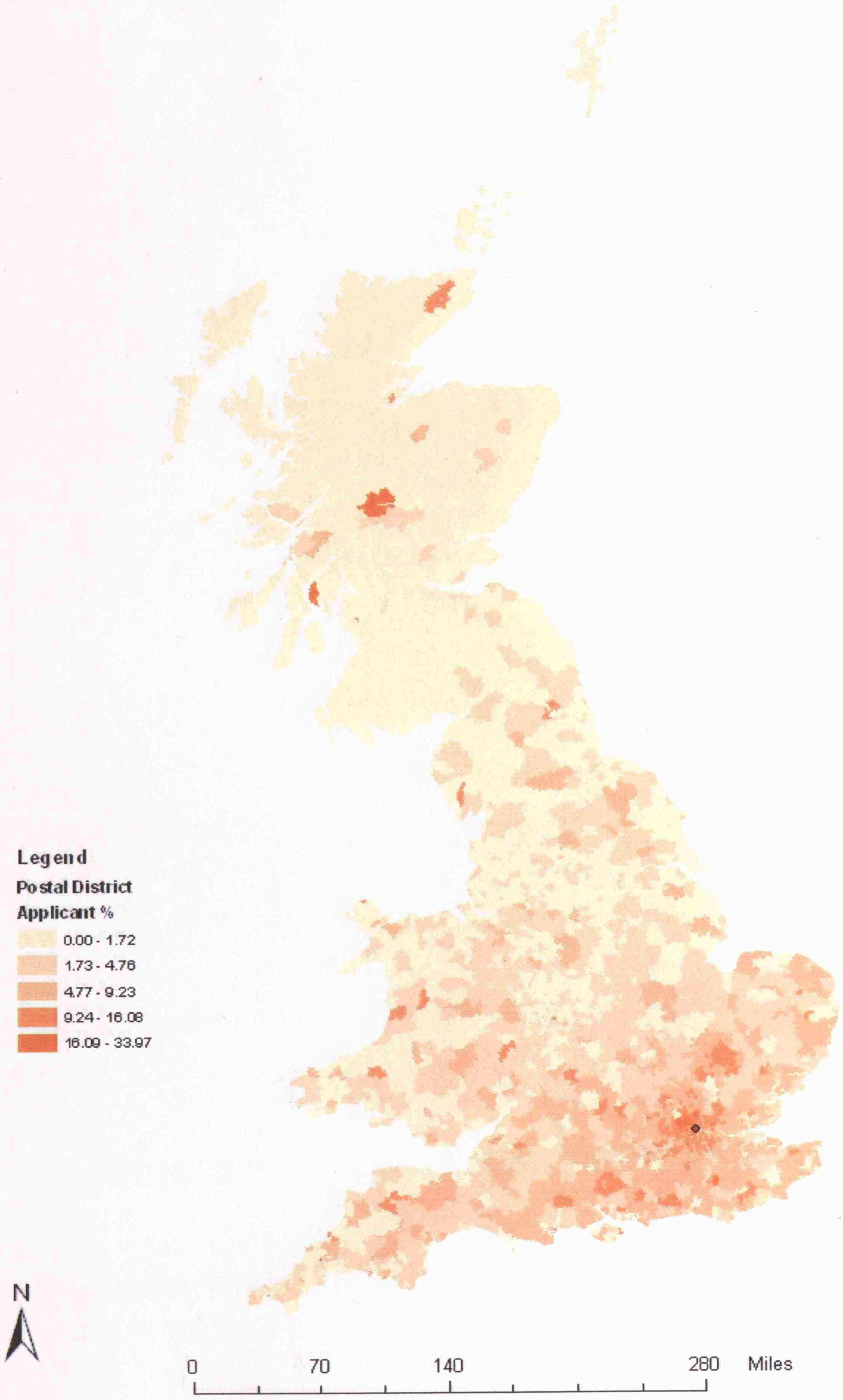


Figure 5.10: UCL 2004 Applicants as a Percentage of all Applicants by Postal Districts

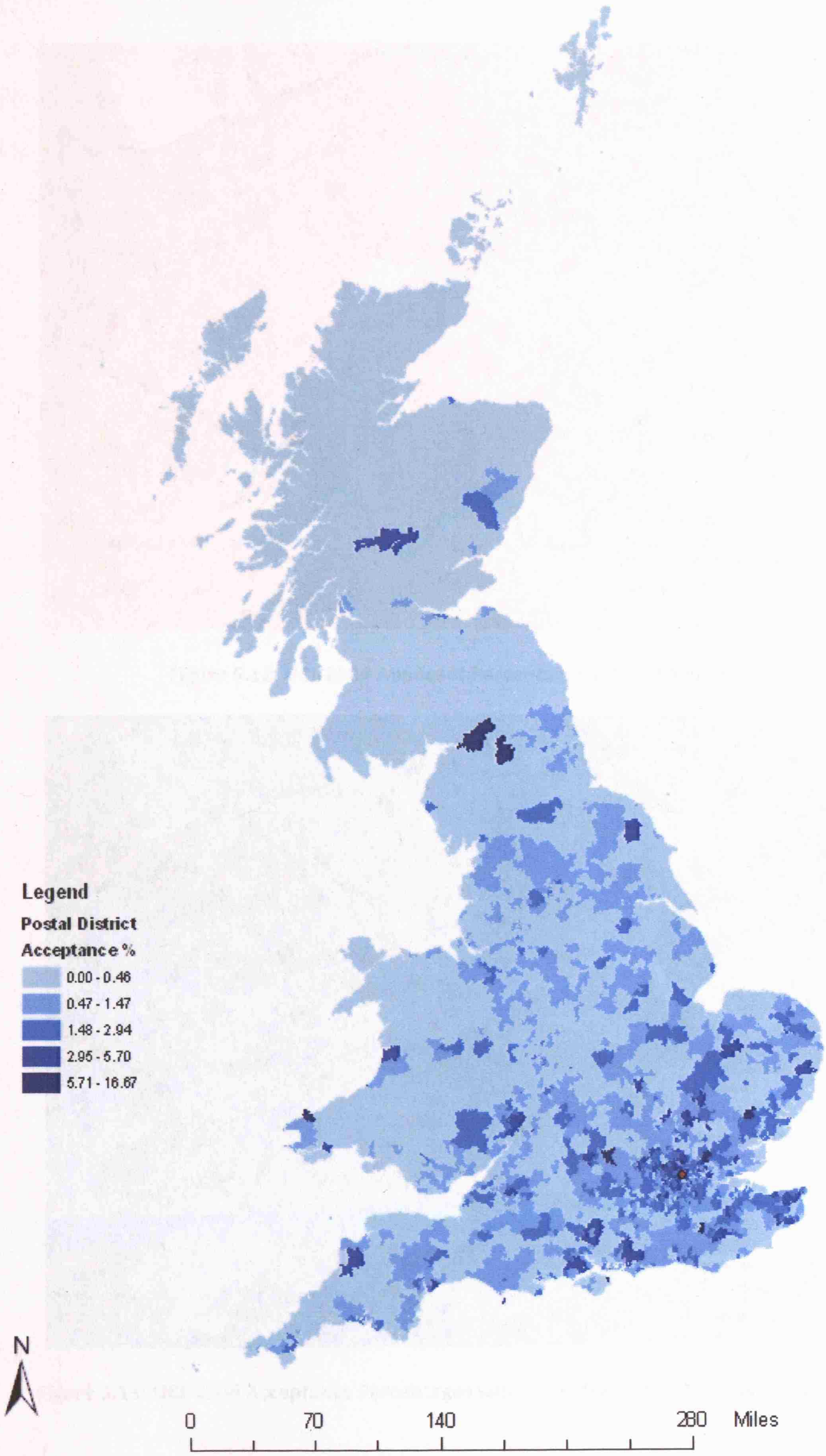


Figure 5.11: UCL 2004 Acceptances as a Percentage of all Acceptances by Postal Districts

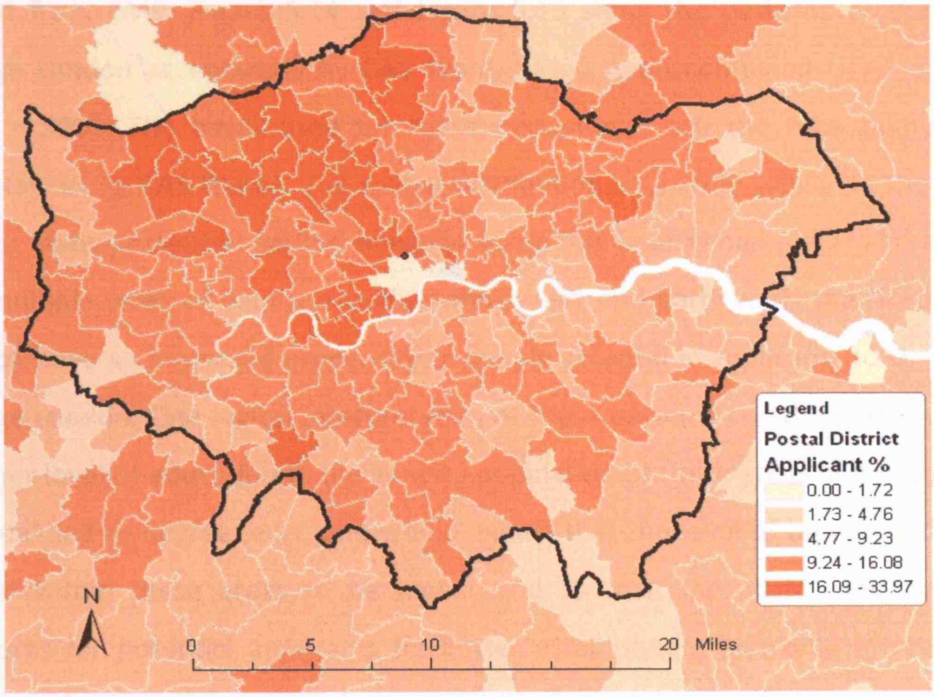


Figure 5.12: UCL 2004 Applicant Percentages within London by Postal Districts.

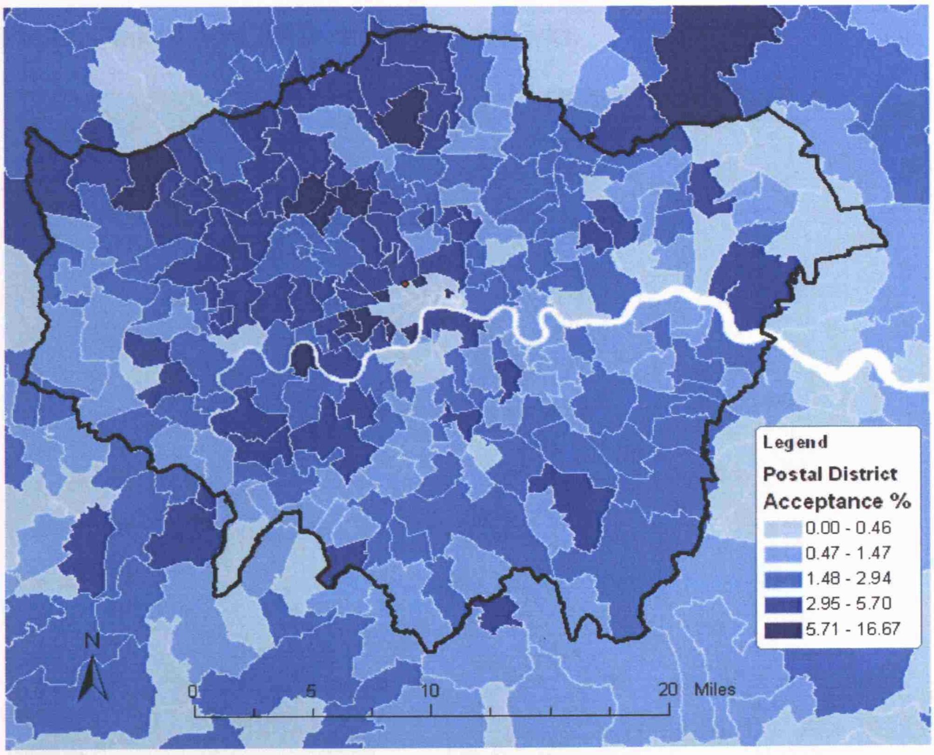


Figure 5.13: UCL 2004 Acceptance Percentages within London (2004 Data) by Postal Districts.

An alternative method of displaying market area using census geography is also output from EMP. Figure 5.14 and Figure 5.15 show the patterns of University College London acceptances by CAS Ward, based upon combined UCAS data for 2002 – 2004. The bases used for these percentages are the total numbers of acceptances by CAS Wards. The spatial distribution is broadly similar to the postal district maps presented earlier, however using this finer geographical aggregation and multiple years of data it is evident that West, and particularly the North West London have an increased propensity to supply acceptable applicants to University College London. One logical interpretation of these patterns could be attributed to a higher level of education in the general population of these areas. If more people are qualified to degree level, one would expect that the level of human and social capital within these areas to be higher and as such the advice and guidance networks for potential applicants be better established. This could increase the chances of acceptance through better targeting by suitable institutions, guidance on interview and application strategy or increased propensity to perform in prior qualifications linked with better support networks.

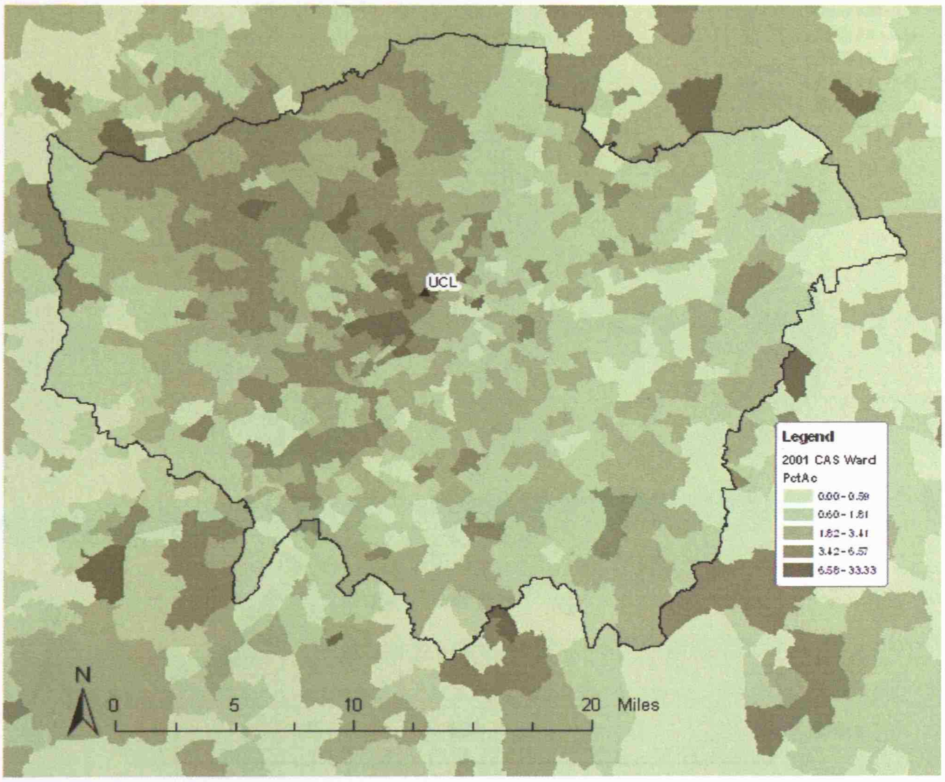


Figure 5.14: UCL Acceptances as a Percentage of all Acceptances by CAS Wards (2002 – 2004 Data) for the London Government Office Region

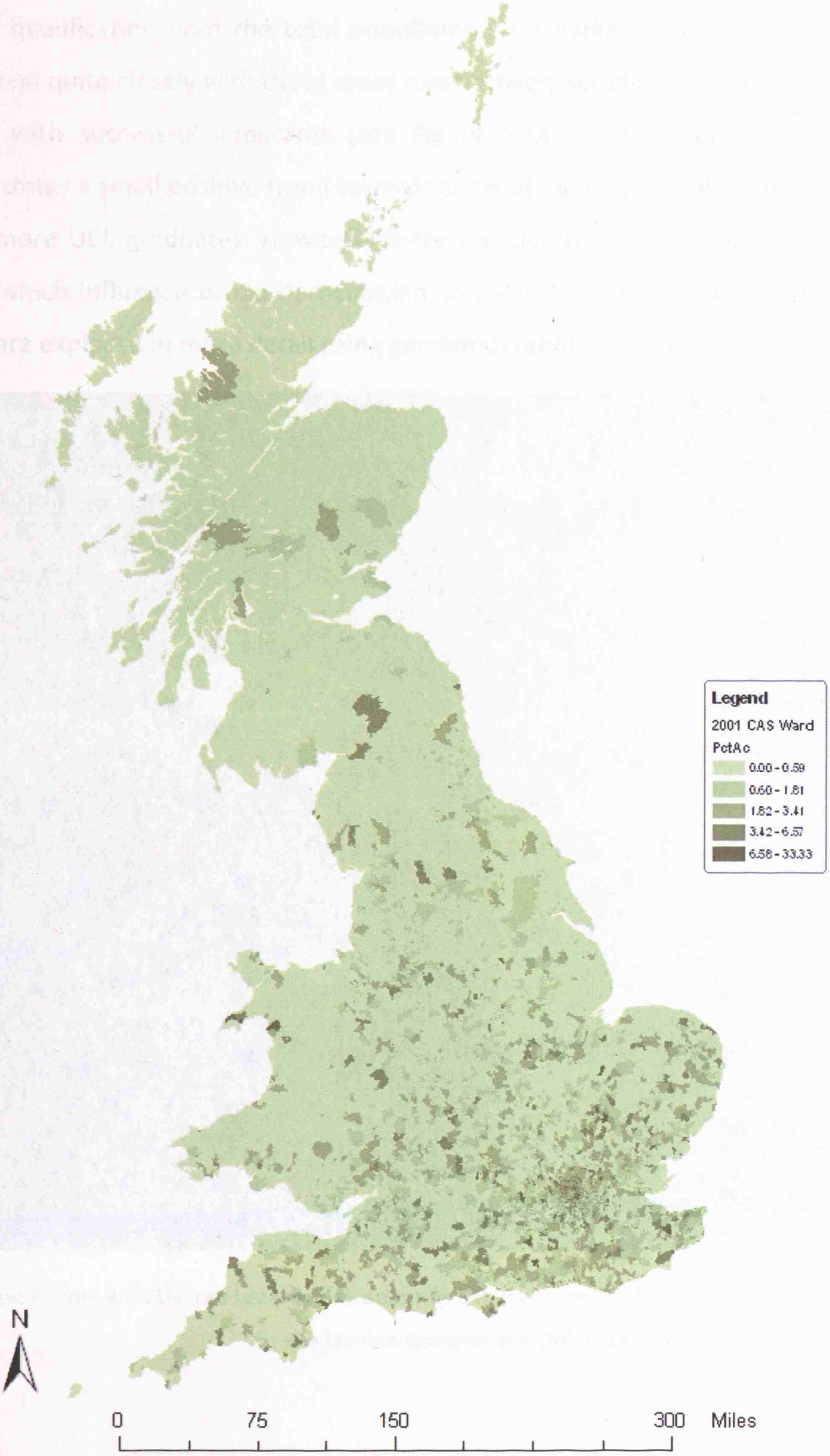


Figure 5.15: UCL Acceptances as a Percentage of all Acceptances by CAS Wards (2002 – 2004 Data)

Figure 5.16 and Figure 5.17 show that there is a relationship between the percentage of people being accepted to University College London and the general level of qualification with the total population. The darker areas on Figure 5.16 correspond quite closely with those areas most actively supplying University College London with successful applicants (see Figure 5.14). Furthermore, Figure 5.17 demonstrates a small positive trend towards areas of higher general qualification to supply more UCL graduates. However, there are clearly a great number of other factors which influence broad participation and also UCL acceptance rates. These factors are explored in more detail using geodemographic analysis.

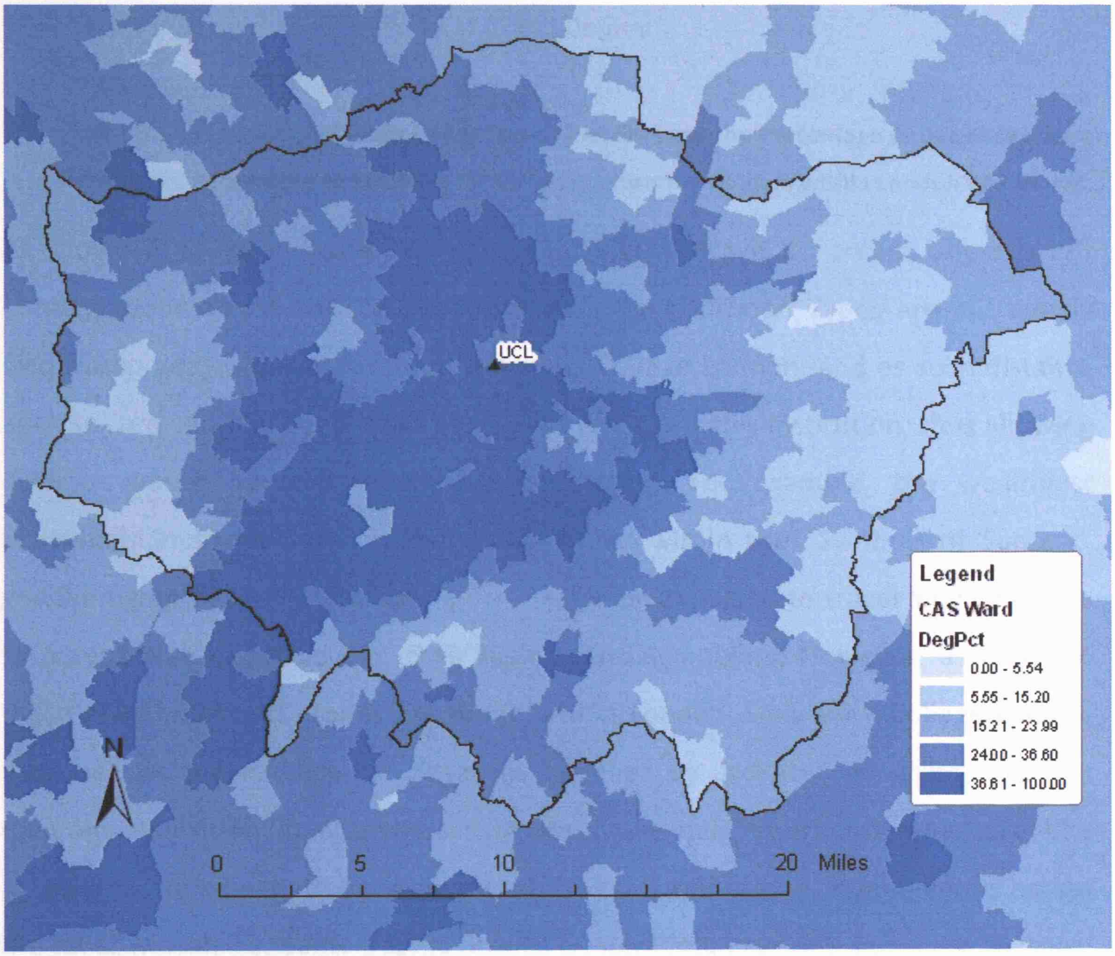


Figure 5.16: People with Degree Level qualifications as a Percentage of all people by CAS Wards for the London Government Office Region (Source: 2001 Census)

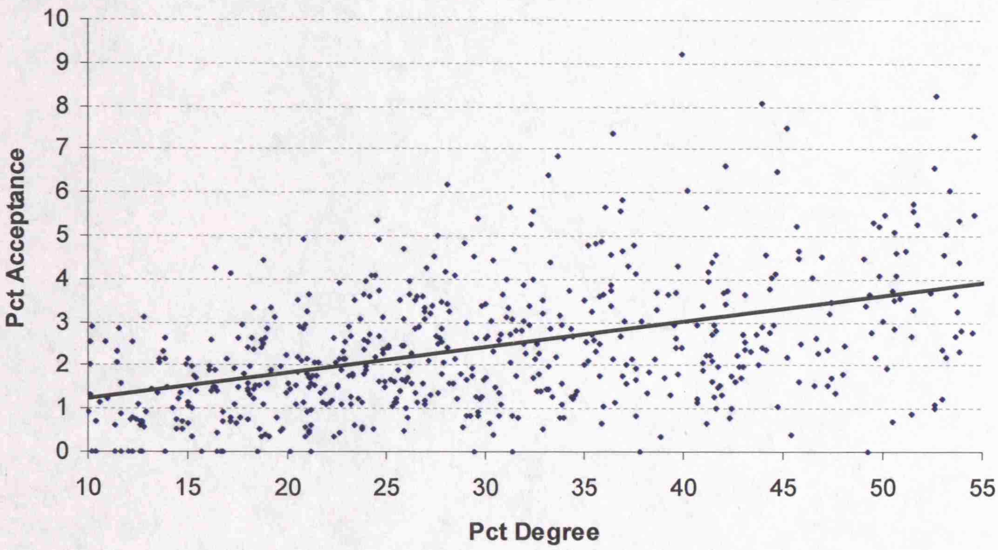


Figure 5.17: A Scatterplot (with Linear Trend Line) Showing the Percentage of UCL Acceptance Against the Percentage of Total Degree Level Qualification Holders within London CAS Wards.

As shown in Chapter 4 the extent that individuals travel to accept offers is not homogeneous across society. Furthermore, neighbourhood Types are not evenly distributed geographically around University College London, and as such distance analyses are sensibly conducted using the profiles for this institution, thus allowing distance based benchmarks to be established. For example, the wealthiest individuals in society are predominantly found within the "Symbols of Success" geodemographic group, and would on average be expected to travel long distances to accept places, because of their high financial mobility. However, as many of these neighbourhood groups are found within London, University College London may logically expect that the distances travelled by accepted applicants are less than an institution in Manchester where these neighbourhood types are less prevalent. The aggregate behavioural effects on travel are not homogenous across the UK as Mosaic geodemographic Groups cluster within regions. For these reasons, distance analysis between geodemographic groups was defined by the University College London acceptance profile, rather than a national average, therefore accounting for the regional differences in the distribution of neighbourhood types. The median average distance that applicants travelled to accept a place at University College London are presented in Table 5.1 and show a distinctive geodemographic travel pattern. This may be related to those neighbourhoods typically found in and around London (e.g. the higher incidence of "Symbols of Success" in the South East and London). When the Mosaic Groups are ranked in descending order of average distance travelled (see Table 5.2), accepted applicants from "Municipal Dependency" neighbourhoods are shown fifth (above "Symbols of Success"), thus travelling longer distances to accept places at University College London, despite these areas containing some of the most deprived neighbourhoods in the UK. There are a number of interpretations of why this pattern may exist. The low incidence of these areas in and around London means that applicants would on average have to travel further to accept a place at University College London. Low acceptance numbers from this group result (small sample size) may affect the accuracy of the averages (see the standard deviation scores in Table 5.3, and the frequencies between years in Table 5.1). In low participation neighbourhoods such as "Municipal Dependency", it is likely that there are specific reasons why individual applicants are attending university, and as such, these unknown factors could affect distance travelled to an institution. For example, a teacher may encourage an individual applicant to University College London, resulting in behaviour not typical of the group averages within these neighbourhoods. This relates to those arguments presented by Fotheringham (1997:88) that "when analysing spatial data, whether by a simple univariate analysis or a more complex multivariate analysis, it may be incorrect to assume that the results obtained from the whole data set apply equally to all parts of the study area".

Table 5.1: Median Distance in Miles by Mosaic Groups that Applicants Travel to Accept Offers, and Numbers of Acceptances at University College London, 2002-2004

	Acceptance Distance			Acceptance Frequency.		
	2002	**2003**	**2004**	**2002**	**2003**	**2004**
Symbols of Success	20.16	18.03	16.09	944	959	878
Happy Families	54.87	42.24	44.12	145	150	132
Suburban Comfort	11.59	11.33	12.70	542	551	542
Ties of Community	9.99	9.59	8.23	272	227	246
Urban Intelligence	6.22	5.49	6.30	328	320	286
Welfare Borderline	4.04	4.27	4.26	103	112	111
Municipal Dependency	18.33	13.04	70.47	15	15	11
Blue Collar Enterprise	26.89	16.11	30.01	74	72	68
Twilight Subsistence	55.43	63.14	41.86	17	15	14
Grey Perspectives	82.27	64.55	55.11	113	118	135
Rural Isolation	113.04	91.75	93.69	172	149	148

Table 5.2: Median Distance in Miles and Mean Acceptances by Mosaic Groups for 2002 – 2004 Acceptances to University College London

	Rank	Average Distance	Average No. Accepted
Rural Isolation	1	99.49	156
Grey Perspectives	2	67.31	122
Twilight Subsistence	3	53.48	15
Happy Families	4	47.08	142
Municipal Dependency	5	33.95	14
Blue Collar Enterprise	6	24.33	71
Symbols of Success	7	18.09	927
Suburban Comfort	8	11.87	545
Ties of Community	9	9.27	248
Urban Intelligence	10	6.00	311
Welfare Borderline	11	4.19	109

"Grey Perspectives" neighbourhoods are ranked second in Table 5.2 and represent those areas where affluent older people still have children attending Higher Education. The increased propensity for these areas to be found in coastal regions, and also have higher disposable income amongst residents could explain why students from these neighbourhoods have increased propensity to travel to University College London. Another interesting finding is the lower rate of travel for applicants from the "Symbols of Success" neighbourhoods. These neighbourhoods have very high income levels, are highly mobile and contain applicants with high levels of attainment. The low distances travelled could be attributed to the prevalence of these areas within London, applicant familiarity with London, the reputation of University College London as a "top university" (Johnson, 2007) and finally the applicant's financial ability to study within the capital. "Welfare Borderline", one of the most underrepresented neighbourhood groups in Higher Education has the lowest propensity to travel to University College London.

Table 5.3: 2002-2004 Median Distance in Miles by Mosaic Types

	Acceptance Distance			Acceptances			Applicants			Standard Deviation *		
	2002	2003	2004	2002	2003	2004	2002	2003	2004	Distance	App.	Acc
Global Connections	2.74	2.81	2.93	92	110	109	531	547	569	0.10	19.08	10.12
Cultural Leadership	7.72	8.03	7.97	201	196	202	1003	999	1033	0.17	18.58	3.21
Corporate Chieftains	16.05	15.18	14.58	269	266	201	1278	1259	1199	0.74	41.24	38.42
Golden Empty Nesters	55.56	51.95	51.87	90	84	97	518	487	513	2.11	16.64	6.51
Provincial Privilege	49.51	57.17	23.80	105	90	77	571	568	518	17.48	29.77	14.01
High Technologists	40.04	31.33	33.97	97	97	103	691	616	616	4.47	43.30	3.46
Semi-Rural Seclusion	53.31	48.67	39.18	90	116	89	634	591	662	7.20	35.76	15.31
Just Moving In	8.04	13.36	51.19	5	22	8	39	94	98	23.53	32.97	9.07
Fledgling Nurseries	36.05	38.89	43.29	17	10	11	84	93	81	3.65	6.24	3.79
Upscale New Owners	67.50	36.74	58.42	32	32	31	208	214	223	15.81	7.55	0.58
Families Making Good	60.03	57.30	37.45	34	31	37	224	265	245	12.32	20.50	3.00
Middle Rung Families	52.18	31.84	35.01	48	40	38	329	298	293	10.94	19.50	5.29
Burdened Optimists	35.09	59.08	23.20	8	13	6	89	101	99	18.28	6.43	3.61
In Military Quarters	86.40	69.57	91.09	1	2	1	10	13	16	11.31	3.00	0.58
Close to Retirement	54.55	37.82	61.32	83	61	69	569	526	504	12.10	33.06	11.14
Conservative Values	62.42	67.71	105.19	20	25	20	185	154	171	23.32	15.52	2.89
Small Time Business	111.68	117.14	104.98	37	47	66	309	299	310	6.09	6.08	14.73
Sprawling Subtopia	17.88	15.23	21.69	51	44	58	384	374	403	3.25	14.73	7.00
Original Suburbs	10.72	10.84	11.30	160	175	170	1015	1075	1105	0.30	45.83	7.64
Asian Enterprise	8.96	9.01	9.39	191	199	159	1266	1198	1209	0.23	36.50	21.17
Respectable Rows	37.75	24.71	67.08	36	42	37	282	275	289	21.70	7.00	3.21
Affluent Blue Collar	97.11	149.63	120.89	24	19	13	148	128	123	26.30	13.23	5.51
Industrial Grit	117.56	59.42	142.91	18	11	13	128	125	109	42.81	10.21	3.61
Coronation Street	152.10	102.92	145.81	8	4	6	61	54	63	26.77	4.73	2.00
Town Centre Refuge	95.55	100.18	67.61	17	14	4	87	65	52	17.62	17.69	6.81
South Asian Industry	7.79	8.25	8.16	38	29	34	309	294	281	0.24	14.01	4.51
Settled Minorities	7.13	6.88	6.90	131	108	139	987	984	983	0.14	2.08	16.09
Counter Cultural Mix	3.75	3.53	3.45	98	99	86	625	621	611	0.16	7.21	7.23
City Adventurers	4.41	4.16	4.51	66	60	49	360	353	343	0.18	8.54	8.62
New Urban Colonists	6.98	6.48	7.81	97	107	98	572	563	570	0.68	4.73	5.51
Caring Professionals	96.64	91.78	98.84	28	24	21	183	190	172	3.61	9.07	3.51
Dinky Developments	17.98	17.16	26.86	13	12	13	91	75	82	5.38	8.02	0.58
Town Gown Transition	107.34	106.10	49.10	18	15	14	101	106	123	33.27	11.53	2.08
University Challenge	95.87	49.73	68.44	8	3	5	40	49	35	23.21	7.09	2.52
Bedsit Beneficiaries	48.52	45.34	39.19	5	9	6	26	41	44	4.74	9.64	2.08
Metro Multiculture	3.86	4.02	3.94	88	90	97	826	869	886	0.08	30.92	4.73
Upper Floor Families	173.26	60.90	75.51	3	5	6	51	45	50	61.10	3.21	1.53
Tower Block Living	0.00	173.52	0.00	0	2	0	13	21	12	100.18	4.93	1.15
Dignified Dependency	2.71	37.19	5.53	7	6	2	36	33	25	19.14	5.69	2.65
Sharing a Staircase	0.00	0.00	0.00	0	0	0	4	5	8	0.00	2.08	0.00
Families on Benefits	10.91	10.47	57.48	7	11	6	67	71	52	27.02	10.02	2.65
Low Horizons	155.28	242.45	57.11	5	2	2	37	34	38	92.72	2.08	1.73
Ex-industrial Legacy	18.33	164.31	146.13	3	2	3	33	28	25	79.56	4.04	0.58
Rustbelt Resilience	111.11	295.63	135.63	9	4	3	44	45	54	100.21	5.51	3.21
Older Right to Buy	97.54	112.08	152.08	4	3	10	39	50	56	28.24	8.62	3.79
White Van Culture	20.61	15.53	15.33	51	54	40	323	347	354	2.99	16.26	7.37
New Town Materialism	44.43	34.91	30.29	10	11	15	111	107	105	7.21	3.06	2.65
Old People in Flats	0.00	0.00	0.00	0	0	0	4	6	4	0.00	1.15	0.00
Low Income Elderly	59.96	90.02	41.86	16	12	14	98	101	100	24.32	1.53	2.00
Cared for Pensioners	36.27	36.27	0.00	1	3	0	5	13	6	20.94	4.36	1.53
Sepia Memories	112.01	56.08	97.25	4	3	3	25	22	22	28.98	1.73	0.58
Childfree Serenity	49.85	18.54	36.11	43	54	65	296	287	294	15.69	4.73	11.00
High Spending Elders	80.02	97.86	86.24	33	32	33	250	198	250	9.05	30.02	0.58
Bungalow Retirement	89.37	63.68	50.80	5	4	3	37	28	35	19.63	4.73	1.00
Small Town Seniors	154.64	107.95	73.76	25	22	28	180	194	169	40.60	12.53	3.00
Tourist Attendants	95.84	171.94	70.49	3	3	3	28	29	17	52.80	6.66	0.00
Summer Playgrounds	228.00	208.81	173.75	10	6	3	39	29	21	27.51	9.02	3.51
Greenbelt Guardians	80.48	66.43	70.78	87	85	87	597	558	560	7.20	21.96	1.15
Parochial Villagers	91.94	95.56	101.25	25	25	24	190	190	201	4.69	6.35	0.58
Pastoral Symphony	150.32	135.23	122.69	41	25	28	214	190	212	13.83	13.32	8.50
Upland Hill Farmers	176.85	170.29	199.48	9	8	6	55	73	48	15.31	12.90	1.53

To illustrate how the spatial distribution of acceptances relates to neighbourhood affluence (associated with other variables such as prior achievement / human, social, economic, cultural capital), the distribution of neighbourhood Types with over representation of people paying the highest tax rate band are averaged across London Census Output Areas (see Table 5.4). A limitation of this method is that Types will not supply equal proportions of the relevant age cohort, and as in the previous analysis, one Type from the 'Grey Perspectives' group is overrepresented. Neighbourhood types were ranked by the index scores for the higher tax rate band (derived from Experian data) and the top 14 were selected. This obviously does not account for the national distribution of these types, so for example "Greenbelt Guardian" neighbourhoods are unlikely to be found within London.

Table 5.4: Mosaic Types Ranked by Higher Tax Rate Bands (Source: Experian)

Rank	Type	Index Higher Rate Tax
1	*A03 Corporate Chieftains*	550
2	*E30 New Urban Colonists*	538
3	*A02 Cultural Leadership*	524
4	*A01 Global Connections*	500
5	*B10 Upscale New Owners*	466
6	*E29 City Adventurers*	426
7	*K58 Greenbelt Guardians*	385
8	*C19 Original Suburbs*	352
9	*A07 Semi-Rural Seclusion*	349
10	*A05 Provincial Privilege*	329
11	*A04 Golden Empty Nesters*	265
12	*A06 High Technologists*	235
13	*J52 Childfree Serenity*	218
14	*E28 Counter Cultural Mix*	214

When the percentages at Output Area for these 14 types are mapped (See Figure 5.18) it can be seen that these patterns are similar to the applicant and acceptance patterns in Figure 5.12 and Figure 5.13.

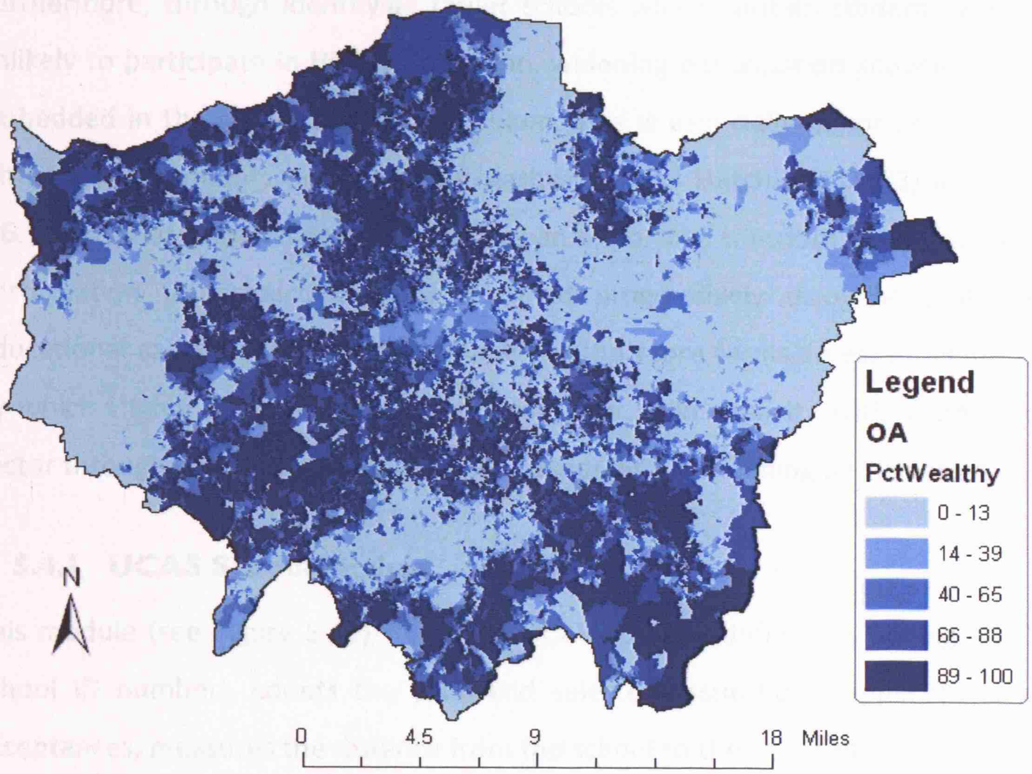


Figure 5.18: Percentage of "Wealthy" Neighbourhoods by 2001 Census Output Areas (Source: Experian)

5.4 Educational Market Profiler for Schools Analysis

The integration of student geodemographic profiles (as defined by Mosaic) for schools with application and acceptance data from UCAS creates a tool allowing institutions to target and benchmark recruitment performance against a potential pool of students who are eligible for university. This analysis enables an institution to target resources more effectively (either for marketing or widening participation) based on the types of students present in schools with defined proximity to the Higher Education institution. As discussed in Section 3.5, the Tooley (1997) concepts of "capacity" and "inclination" vary between neighbourhood Groups and provide a theoretical explanation as to why differentiation in aggregate participation behaviours occur between neighbourhood groups as recorded in the empirical analysis in Chapter 4. Targeting at a school level rather than directly to individuals is beneficial to an institution as it maximises potential return from a single investment in an event, i.e. reaching more of an intended target with a single initiative.

Furthermore, through identifying target schools which contain students who are unlikely to participate in Higher Education, widening participation activities can be embedded in the students prior curriculum. This is especially important and links with those arguments presented by Leathwood and Hutchings (2003) in Section 2.6.1 that poor prior attainment, which is an important contributory factor to non participation is socioeconomically stratified progressively throughout students educational careers. This school level analysis therefore forms an essential method by which Higher Education institutions can help build capacity within the school sector through strategic targeting of their resources for widening participation.

5.4.1 UCAS School Module

This module (see Figure 5.19) selects all UCAS schools (defined as having a UCAS school ID number), counts the total and selected institution's applications and acceptances, measures the distance from the school to the selected institution, and calculates whether the school falls within the average distance that applicants from each of the 61 Mosaic Types are likely to travel to the selected institution to accept a place. This provides a basic analysis of which schools actively supply students to Higher Education in general and specifically those supplying students to the chosen institution.

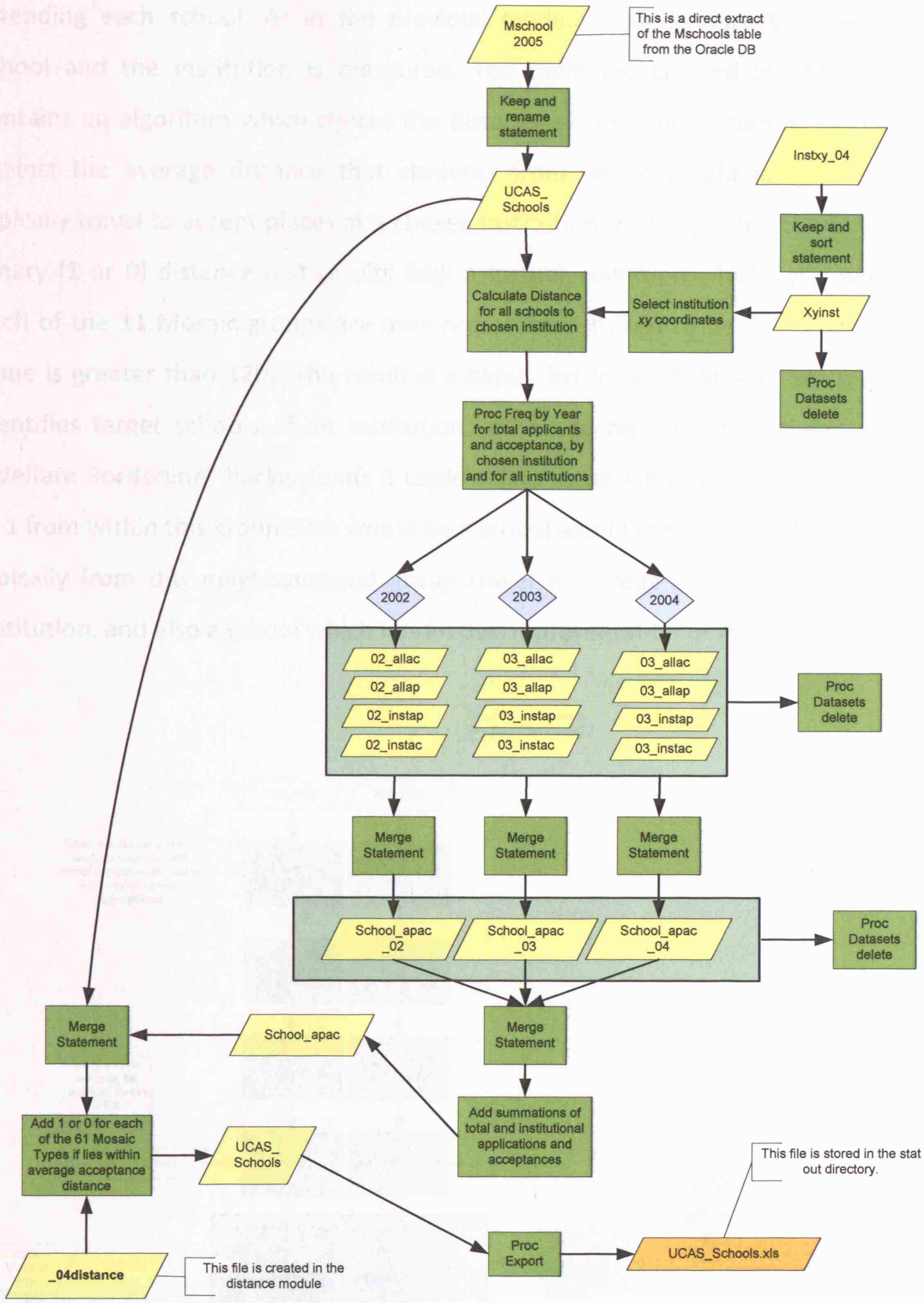


Figure 5.19: UCAS School Data Analysis Module

5.4.2 The PLASC Schools Module

This module (see Figure 5.20) creates index scores for each of the eleven 2001 Mosaic Groups across all secondary schools in England. These index scores are derived using annual PLASC data and as such are based on the actual students attending each school. As in the previous module, the distance between each school and the institution is measured. The Excel file created by this module contains an algorithm which checks the distance of the school from an institution against the average distance that students from within neighbourhood groups typically travel to accept places at a chosen institution. The algorithm combines the binary (1 or 0) distance test results with a further condition which tests whether each of the 11 Mosaic groups are over represented (taken to be where the index value is greater than 120). The result is a binary list for each Mosaic Group which identifies target schools. If an institution wished to recruit more students from "Welfare Borderline" backgrounds it could select those schools with a target score of 1 from within this group. This would be a school within the distance that students typically from this neighbourhood group travel to accept a place at the chosen institution, and also a school which has an overrepresentation of these students.

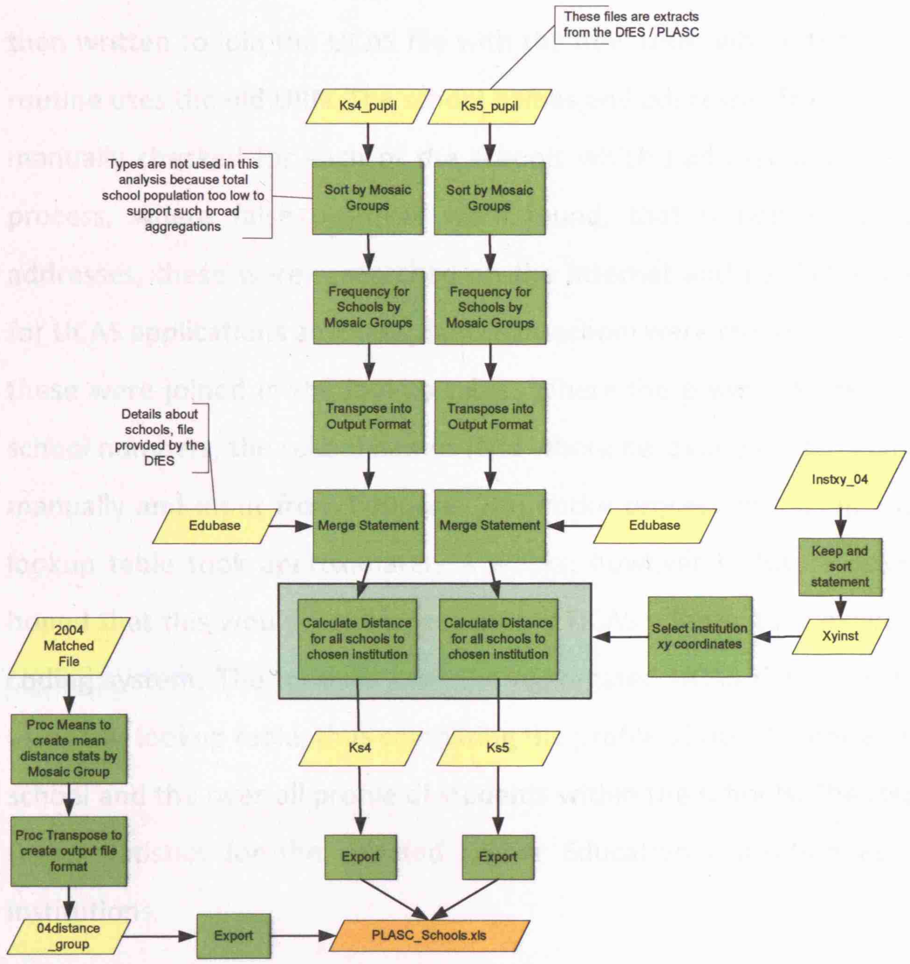


Figure 5.20: PLASC School Data Analysis Module

5.4.3 Stitching Module

The stitching module (see Figure 5.21) creates a very useful output for targeting and also widening access to underrepresented groups. This module aggregates UCAS applicants and acceptances by schools to geodemographic profiles of the students within the total school population. UCAS and the DCSF do not use a single compatible school coding schemes and there is no official lookup table through which matches could be made. Thus before the module was written a lookup table was created to join the DCSF database of schools (Edubase) to the UCAS schools codes. Firstly the LEA code and Establishment code on the Edubase database were concatenated and assigned a unique reference number (URN). Two types of codes existed on this file (old and new). The concatenation was made for all new codes and also older codes where available, which related to a previous coding scheme. Secondly the LEA code and Establishment code on the UCAS schools database were concatenated to form a unique reference number. A two step merge statement was then written to join the UCAS file with the new URN: where there was no join, the routine uses the old URN. The school names and addresses from each file were then manually checked for each of the schools which had matched records. From this process, where false positives were found, that is non matching names and addresses, these were researched on the Internet and corrected by hand. Counts for UCAS applications and acceptances by school were created for 2002 to 2004 and these were joined in the lookup table. Where there was no linking URN for UCAS school numbers, the school names (and where necessary postcodes) were searched manually and input from Edubase. The entire process of joining and checking the lookup table took approximately 4 weeks, however in future developments it is hoped that this would not be necessary if UCAS adopts, as is expected, a standard coding system. The module joins the aggregated UCAS data with the PLASC data using this lookup table, thus comparing the profile of the student accepted from the school and the over-all profile of students within the schools. The routine calculates these statistics for the selected Higher Education institution as well as for all institutions.

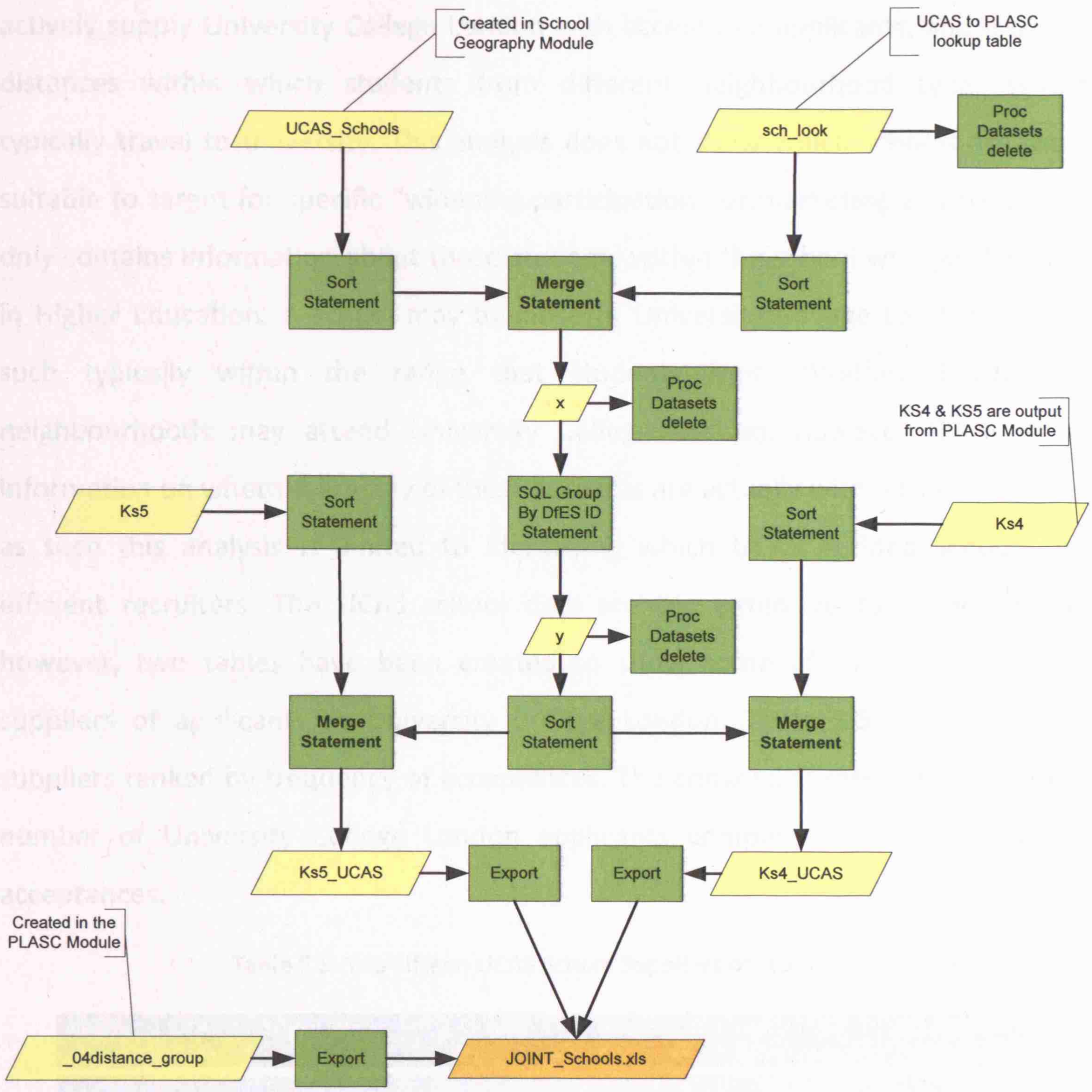


Figure 5.21: School Data Stitching Module

5.4.4 School Analysis Case Study – University College London

The school analysis provided the information necessary for the University College London outreach team to make strategic decisions that would be most efficiently achieved (financially) about which schools to target for extending participation to underrepresented groups or targeting recruitment for specific courses.

The UCAS school module measured the distance between each UCAS defined schools and the average distance that accepted applicants travelled from their homes to University College London. These distances were disaggregated by Mosaic neighbourhood Groups, therefore allowing identification of those schools which actively supply University College London with acceptable applicants, and also the distances within which students from different neighbourhood types would typically travel to university. This analysis does not show which schools are most suitable to target for specific "widening participation" or marketing activities, as it only contains information about those students within the school who participated in Higher Education. A school may be close to University College London, and as such typically within the range that students from "Welfare Borderline" neighbourhoods may attend University College London, however there is no information on whether are any of these students are actually within the school and as such this analysis is limited to identifying which UCAS defined schools are efficient recruiters. The UCAS school data are too extensive to report in full; however, two tables have been created to show some of the most frequent suppliers of applicants to University College London. Table 5.5 shows the top suppliers ranked by frequency of acceptances. The conversion rate is based on the number of University College London applicants compared to the number of acceptances.

Table 5.5: Top Fifteen UCAS School Suppliers of Students to UCL (2002 – 2004)

Rank	UCAS Code	School Name	UCL Applicants	UCL Acceptances	UCL School Conversion
1	11898	Woodhouse College	343	62	18.08
2	10172	Hills Road VI Form College	269	59	21.93
3	12277	Richmond Upon Thames College	382	51	13.35
4	12232	St Dominic's VI Form College	317	48	15.14
5	12251	Merchant Taylor's School	134	44	32.84
6	11935	Latymer School	232	43	18.53
7	10649	City & Islington Sixth Form College	200	40	20.00
8	14379	Mander Portman Woodward	178	38	21.35
9	12060	Westminster School	243	38	15.64
10	11883	City Of London School	168	37	22.02
11	11055	Peter Symonds College	248	34	13.71
12	12904	Tiffin School For Boys	140	33	23.57
13	12019	Dulwich College	151	33	21.85
14	11111	Haberdashers Askes Boys School	183	33	18.03
15	11815	St Paul's School	175	32	18.29

Table 5.6 identifies those schools within 50 miles of UCL that supplied 16 or more applicants over the 2002-2004 period. These are then ranked by the conversion rate. Although these schools do not all send a high volume of accepted applicants

they do appear to have a loyalty to UCL, possibly providing favourable advice and guidance to potential applicants.

Table 5.6: Top Fifteen UCAS School Suppliers of Converted Students to UCL (2002 – 2004) Supplying >15 total Applicants and Within 50 miles of UCL

Rank	UCAS Code	School Name	UCL Applicants	UCL Acceptances	UCL School Conversion	Distance
1	11101	St Margarets School	18	10	55.56	12.63
2	12315	Peniel Academy	23	11	47.83	19.97
3	10062	Abbey School	48	19	39.58	36.32
4	12869	Farnham College	18	7	38.89	36.14
5	11146	Royal Masonic School	19	7	36.84	16.98
6	10645	Chigwell School	71	26	36.62	11.68
7	12227	Heathfield School	38	13	34.21	11.60
8	12882	Guildford High School For Girls	47	16	34.04	27.17
9	12254	St Helens School	62	21	33.87	13.93
10	11958	Hasmonean High School (Girls)	83	28	33.73	6.97
11	11164	St Columbas College	21	7	33.33	17.84
12	16267	Wentworth Tutorial College	18	6	33.33	5.17
13	12251	Merchant Taylors School	134	44	32.84	14.53
14	11322	Maidstone Grammar School For Girls	34	11	32.35	32.86
15	12076	Putney High School (G.D.S.T)	47	15	31.91	5.89

As mentioned above, the UCAS data on schools alone cannot show an institution the aggregate profile of all students within a school. Thus, the PLASC School Module created geodemographic profiles for students within schools at Key Stages 4 and 5 (GCSE & A-Level). Using these outputs the schools were ranked by their neighbourhood Group composition. Using this information University College London could target particular neighbourhood Groups for widening participation or marketing activities at both Key Stages 4 & 5. A number of tables are provided to illustrate some of these outputs which rank the most overrepresented neighbourhood Groups within the average distances that these groups travel to accept places at University College London (See Table 5.7 - Table 5.9).

Table 5.7: Key Stage 4 Highest Propensity for "Urban Intelligence" Students within 24.20 miles of UCL (Source: 2004 UCAS Data)

Rank	School Name	LEA	Establishment No.	Index Score
1	William Morris Academy	205	4320	3136
2	John Evelyn Education Centre	209	1101	3136
3	Queen Elizabeth II Jubilee School	213	7184	2352
4	Paddock School	212	7183	2090
5	The Bridge School	206	7031	2090
6	Samuel Rhodes MLD School	206	7146	1960
7	College Park School	213	7042	1881
8	Jack Tizard School	205	7203	1881
9	Acland Burghley School	202	4285	1748
10	Stoke Newington School	204	4310	1698
11	The Camden School for Girls	202	4611	1612
12	Hampstead School	202	4275	1599
13	Jack Taylor School	202	7185	1568
14	Richard Cloudesley PH School	206	7030	1568
15	Lansdowne School	208	7001	1568

Table 5.8: Key Stage 4 Highest Propensity for "Grey Perspectives" Students within 74.73 miles of UCL (Source: 2004 UCAS Data)

Rank	School Name	LEA	Establishment No.	Index Score
1	The Leys School	873	6003	2321
2	Belstead School	935	7005	1160
3	Manhood Community College	938	4037	1148
4	Clacton County High School	881	5444	1139
5	Seaford Head Community College	845	4036	1015
6	Willingdon Community School	845	4039	980
7	Tendring Technology College	881	5432	971
8	Highfield Special School	873	7007	928
9	Bexhill High School	845	4044	801
10	St Richard's Catholic College	845	4606	788
11	Maplewood School	825	7000	774
12	Swalcliffe Park School Trust	931	7007	774
13	Ursuline College	886	4633	750
14	The Angmering School	938	4060	749
15	Chatham House Grammar School for Boys	886	5462	737

Table 5.9: Key Stage 5 Highest Propensity for "Welfare Borderline" Students within 15.16 miles of UCL (Source: 2004 UCAS Data)

Rank	School Name	LEA	Establishment No.	Index Score
1	City and Islington College	206	8033	2293
2	Ashbourne Independent School[39]	207	6348	2293
3	Lewisham College	209	8006	2293
4	Tower Hamlets College	211	8066	2293
5	Stanmore College	310	8002	2293
6	Archbishop Tenison's School	208	5403	1681
7	Archbishop Michael Ramsey Technology College	210	4725	1681
8	Deptford Green School	209	4047	1681
9	South Camden Community School	202	4196	1660
10	Gladesmore Community School	309	4033	1528
11	St Aloysius RC College	206	4651	1528
12	Addey and Stanhope School	209	4600	1528
13	Charles Edward Brooke School	208	4509	1490
14	St Saviour's and St Olave's Church of Eng	210	4680	1433
15	Haverstock School	202	4104	1386

By combining the 2002-2004 UCAS and 2004 PLASC schools data using the Stitching Module a powerful tool was created to help apportion limited "widening participation" resources most efficiently. This tool allowed University College London to select those "target" schools with a high frequency of students from "Welfare Borderline" neighbourhoods, already discussed as the most underrepresented in Higher Education (See Section 4.4.3), and assess the propensity for these schools to supply applicants to University College London. Figure 5.22 shows the spatial distribution of the schools which were classified as a target (in red) with the size of the circles increasing relative to the frequency of University College London acceptances 2002-2004. The schools were overlaid on top of index of multiple deprivation scores by Super Output Areas to illustrate that most schools high in students from "Welfare Borderline" areas are situated in more generally deprived areas. Target schools had index scores for "Welfare Borderline" over 120 and were also within the 15.16 miles of University College London

[39] Ashbourne Independent School has a single student recorded on the PLASC data for KS5. This student is from a "Welfare Borderline" neighbourhood, and as such explains their presence in the table. Independent schools are not obliged to submit PLASC returns, and as such their data may be missing or incomplete.

(average mean distance that University College London "Welfare Borderline" applicants travel to accept a place).

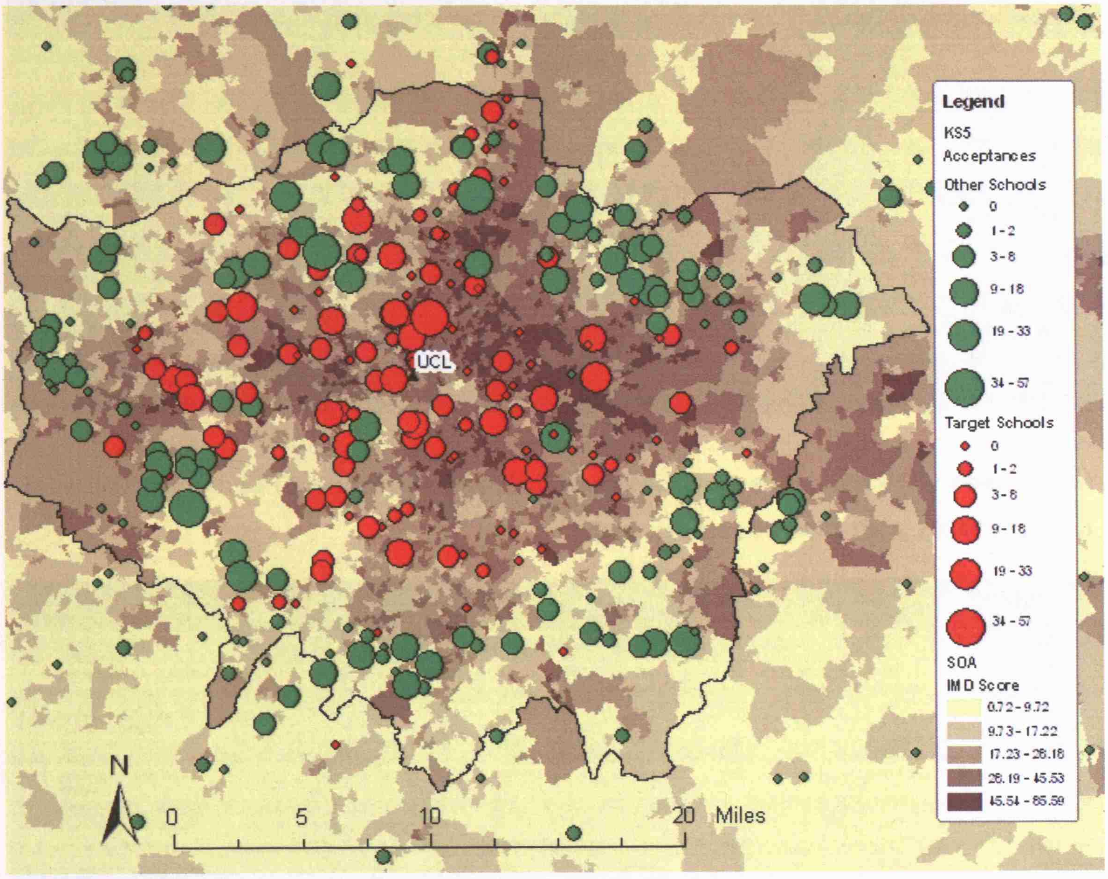


Figure 5.22: Target Schools and the Index of Multiple Deprivation as 5 Jenks

To support this map several tables were created to rank the target schools from which UCL recruited the most students (based on 2002-2004 UCAS data: Table 5.10) and the target schools from which UCL recruits the fewest students (Table 5.11).

Table 5.10: Top 15 Target Schools ranked by Acceptance Frequency (Source: 2002-04 UCAS Data, 2004 DfES Data)

Rank	School Name	Welfare Borderline Frequency	Distance	LEA	DfES Estab. Code	Welfare Borderline Index	02 - 04 UCL Applicants	02 - 04 UCL Acceptances	02 - 04 All Applicants	02 - 04 Total Acceptances	UCL Conversion %	Total Conversion %	UCL Penetration %
1	City and Islington College	1	2.32	206	8033	2293	377	53	2896	2190	14.06	75.62	2.42
2	Preston Manor High School	9	7.28	304	5410	211	92	25	301	280	27.17	93.02	8.93
3	St Michael's Catholic Grammar School	7	6.54	302	5404	137	87	20	333	304	22.99	91.29	6.58
4	Newham Sixth Form College	1	7.22	316	8600	1146	170	19	1673	1361	11.18	81.35	1.40
5	William Morris Academy	53	4.34	205	4320	611	117	18	629	523	15.38	83.15	3.44
6	The London Oratory School	23	3.79	205	5400	340	106	18	486	412	16.98	84.77	4.37
7	The Cardinal Vaughan Memorial RC School	22	3.64	207	5402	420	84	17	360	327	20.24	90.83	5.20
8	St Angela's Ursuline Convent School	23	7.23	316	4600	533	112	16	600	519	14.29	86.50	3.08
9	La Sainte Union Catholic Secondary School	34	2.53	202	5401	757	60	15	283	243	25.00	85.87	6.17
10	The Grey Coat Hospital	20	2.00	213	4628	546	78	15	243	209	19.23	86.01	7.18
11	Haberdashers' Aske's Hatcham College	24	5.60	209	6900	437	58	15	379	325	25.86	85.75	4.62
12	The Cardinal Wiseman Roman Catholic Schoo	11	9.42	307	4603	332	48	15	288	250	31.25	86.81	6.00
13	Tower Hamlets College	1	5.24	211	8066	2293	148	12	1288	990	8.11	76.86	1.21
14	Graveney School	16	7.13	212	5400	165	119	12	591	514	10.08	86.97	2.33
15	The St Marylebone CofE School	19	0.78	213	4673	641	63	11	201	173	17.46	86.07	6.36

Table 5.11: Bottom 15 Target Schools ranked by Acceptance Frequency (Ordered by Index Scores) (Source: 2002-04 UCAS Data, 2004 DfES Data)

Rank	School Name	Welfare Borderline Frequency	Distance	LEA	DfES Estab. Code	Welfare Borderline Index	02 - 04 UCL Applicants	02 - 04 UCL Acceptances	02 - 04 All Applicants	02 - 04 Total Acceptances	UCL Conversion %	Total Conversion %	UCL Penetration %
1	Bishop Challoner Catholic Collegiate Girl	33	3.77	211	4726	1376	6	0	154	113	0.00	73.38	0.00
2	Central Foundation Girls' School	24	4.64	211	4507	764	26	0	128	95	0.00	74.22	0.00
3	William Ellis School	24	2.47	202	4688	567	0	0	0	0	0.00	0.00	0.00
4	Archbishop Michael Ramsey Technology College	22	3.59	210	4725	1681	14	0	107	72	0.00	67.29	0.00
5	Cardinal Pole Roman Catholic School	22	4.51	204	4714	1363	20	0	120	91	0.00	75.83	0.00
6	Sir John Cass Foundation and Redcoat	20	3.98	211	4722	1019	15	0	127	90	0.00	70.87	0.00
7	Sedgehill School	14	8.60	209	4267	563	4	0	74	55	0.00	74.32	0.00
8	St Bonaventure's RC School	14	7.02	316	4601	345	0	0	0	0	0.00	0.00	0.00
9	Charles Edward Brooke School	13	3.69	208	4509	1490	7	0	40	24	0.00	60.00	0.00
10	Blackheath Bluecoat Church of England	13	7.81	203	4715	903	1	0	54	47	0.00	87.04	0.00
11	Ernest Bevin College	12	6.23	212	4297	598	13	0	112	84	0.00	75.00	0.00
12	Forest Hill School	12	7.47	209	4289	585	7	0	125	107	0.00	85.60	0.00
13	The Skinners' Company's School for Girls	11	4.26	204	4686	1051	4	0	68	53	0.00	77.94	0.00
14	BRIT School for Performing Arts	10	9.52	306	6900	247	5	0	363	226	0.00	62.26	0.00
15	George Green's School	9	6.07	211	4505	938	13	0	127	92	0.00	72.44	0.00

Table 5.10 and Table 5.11 showed University College London which of the target schools where supplying the least and most accepted applicants respectively. Other useful information provided in the tables were the conversion rates for applicants making applications to Higher Education, and also specifically for University College London. Finally a penetration score was derived by comparing University College London acceptances with the total acceptances in the particular school/ college. Target schools were selected on the basis of having a 'Welfare Borderline' index scores of over 120 (i.e. overrepresented) and within the average distance that students from these neighbourhoods would travel to accept places at University College London. The University College London conversion rate is the percentage of University College London applicants who convert into University College London

acceptances. The total conversion rate is the percentage of total applicants who convert into any institutions acceptances. The University College London penetration score is the percentage of University College London acceptances compared to the total acceptance frequency.

In Table 5.10 which showed those target schools supplying University College London with the most acceptances contained a number of interesting findings. The London Oratory School[40] appeared in this list which is a high performing state school. On further examination of this school it was shown to have a bimodal entry profile, containing overrepresentation of students from "Welfare Borderline" and also "Symbols of Success" neighbourhoods. Because of this dichotomy the acceptance rate must be interpreted cautiously. The applicants from this school who were accepting places were examined in detail and it was found that only four came from the "Welfare Borderline" neighbourhoods, and the other fourteen from more affluent areas. A second finding was that there appeared to be an over representation of Catholic Faith Schools (including the London Oratory School) in both tables. However, if the neighbourhood profiles of Key Stage 5 Roman Catholic faith schools are examined (See Figure 5.23 and Figure 5.24), it is clear that across both England and London, these school types tend to have students from "Welfare Borderline" neighbourhoods, therefore explaining why many of these schools appear in the lists.

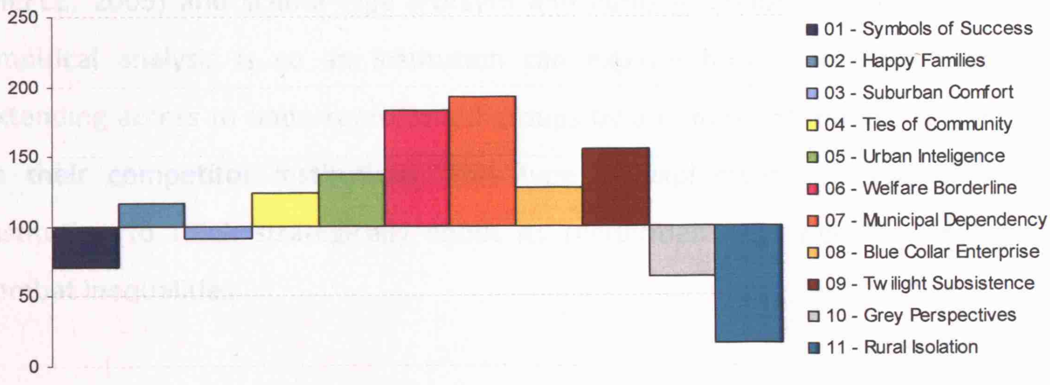


Figure 5.23: KS5 London Roman Catholic School Neighbourhood Profiles (Source: 2004 DCSF Data)

[40] Factoid: Tony Blair sent his children to this school.

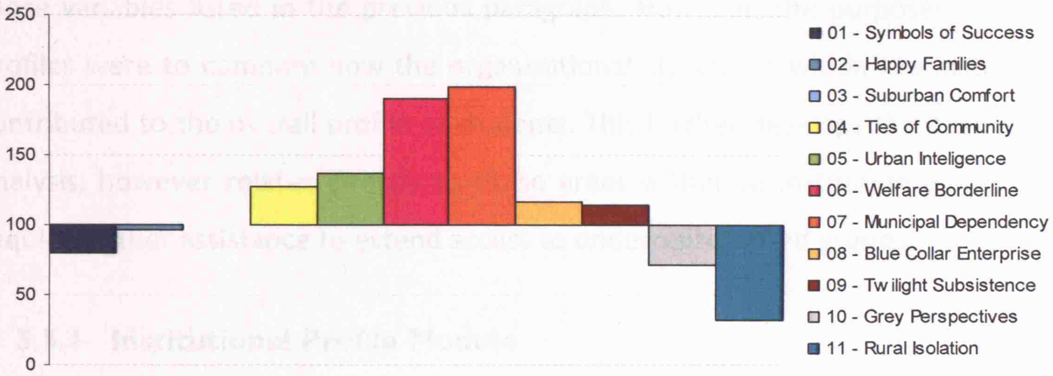


Figure 5.24: KS5 England Roman Catholic School Neighbourhood Profiles (Source: 2004 DCSF Data)

5.5 Educational Market Profiler for Institutional Benchmarking

A large component of EMP creates detailed institutional profiles which compare the recruitment of the students at the institution being analysed with their main competitors (defined by the institution). This external benchmark analysis form a series of cross tabulations of JACS course groups against gender, geodemographics, ethnicity, NS-SEC, POLAR and school type. These analyses therefore show the behaviour of people classified within these groups to participate in different courses of study at a Higher Education institution relative to their competitors. The variables were chosen specifically as they related to literature on access, for example gender (Egerton and Halsey, 1993; Thomas, 1990), geodemographics (see Chapter 4), ethnicity (Modood, 1993, Reay *et al*, 2005), NS-SEC (Archer *et al*, 2003), POLAR (HEFCE, 2005) and school type (Forsyth and Furlong, 2000). The purpose of the empirical analysis is so an institution can explore how their performance at extending access to underrepresented groups by a number of measures compares to their competitor institutions. This type of exploratory analysis allows an institution to think strategically about its recruitment and develop strategy to combat inequalities.

The final component of EMP profiled each department and faculty within an institution against the total institutional profiles. This was also carried out for all those variables listed in the previous paragraph. However, the purposes of these profiles were to compare how the organisational structures within the institution contributed to the overall profile of students. This further develops the descriptive analysis, however relates directly to those areas within an institution which may require further assistance to extend access to underrepresented groups.

5.5.1 Institutional Profile Module

This module (see Figure 5.25) compares the chosen institution entry profile against the national profile, and a competitor group (as defined by a given by institution). The variables used in the comparisons include geodemographic type, gender, ethnicity, socio-economic group, POLAR score and School Type. The module also compares internal department and faculty (course aggregation list provided by institution) aggregations against a base of the total institutional profile. This allows for an assessment of the contribution to the total institutional profile by particular areas within the institution. Finally, course (defined by JACS) profiles are compared with the same aggregations in competitor institutions. The output from each of these analyses is formatted in an Excel template.

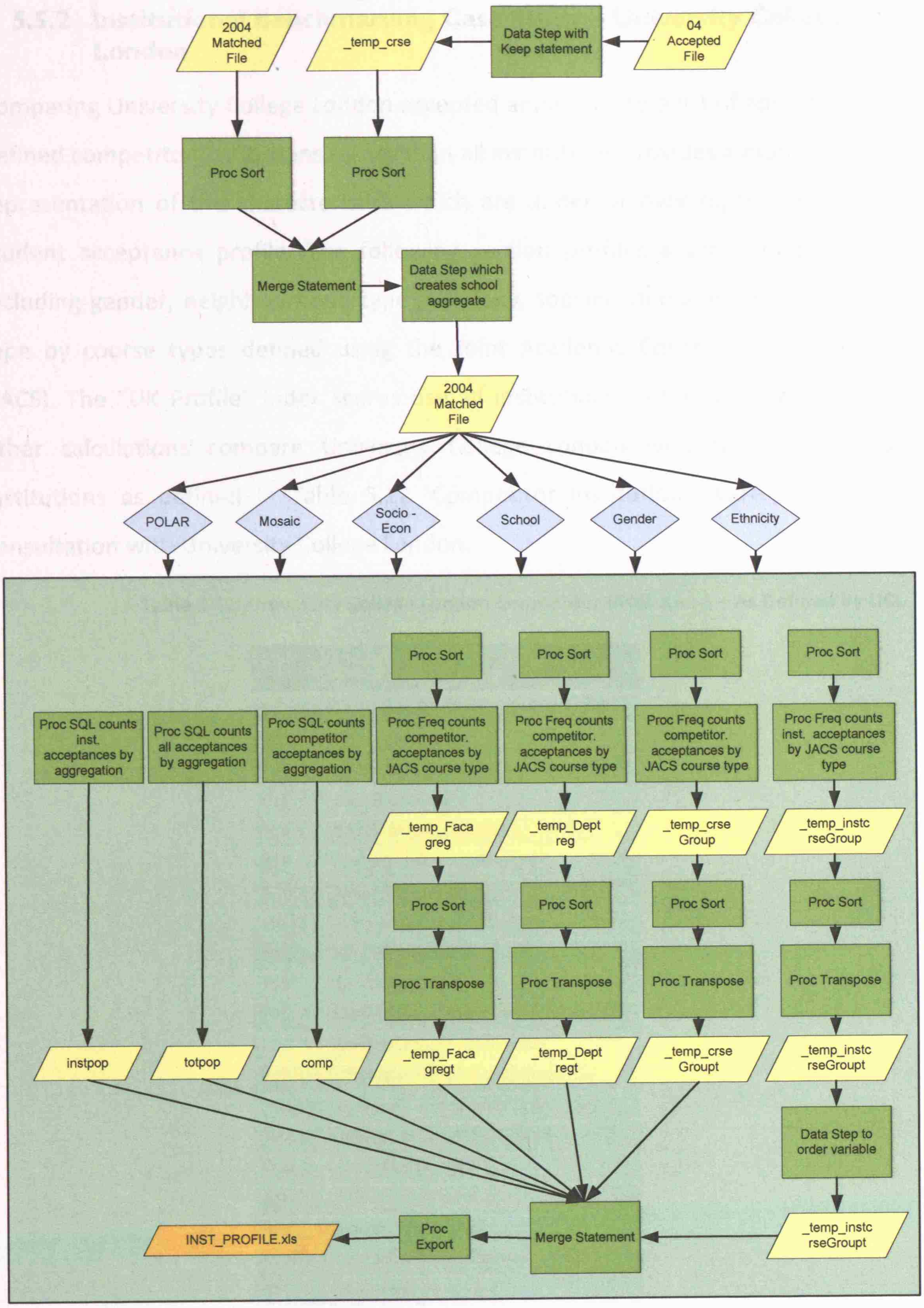


Figure 5.25: Institutional Profile Module

5.5.2 Institutional Benchmarking Case Study – University College London

Comparing University College London accepted applicants to a list of appropriately-defined competitor institutions rather than all institutions provides a more accurate representation of the characteristics which are under or over represented in its student acceptance profile. The following section profiles a series of segments including gender, neighbourhood type, ethnicity, socioeconomic group and school type by course types defined using the Joint Academic Course Coding Scheme (JACS). The "UK Profile" index scores use all institutions as the base, whereas all other calculations compare University College London with their competitor institutions as defined in Table 5.12. Competitor institutions were selected in consultation with University College London.

Table 5.12: University College London Competitor Institutions – As Defined by UCL

Code	Institution
B32	The University of Birmingham
B78	University of Bristol
C05	University of Cambridge
C15	Cardiff University
D86	The University of Durham
E56	The University of Edinburgh
G28	University of Glasgow
I50	Imperial College London
K60	Kings College London
L23	University of Leeds
L41	The University of Liverpool
L72	London School of Economics
M20	The University of Manchester
N21	University of Newcastle Upon Tyne
N84	The University of Nottingham
O33	Oxford University
S18	The University of Sheffield
S27	University of Southampton
W20	The University of Warwick
Y50	The University of York

The profiles are presented as a series of graphs and tables using the 2004 UCAS data. Table 5.13 - Table 5.17 consist of index scores and a "traffic light system" to aid interpretation. A Green circle indicates an index score >= 120, yellow between 80 and 119, and red <= 80.

Table 5.13: University College London Gender Profiles

Profile	*Male*	*Female*
UK Profile	**105**	**95**
Competitor Profile	**102**	**98**
A Medicine & Dentistry	119	87
B Subjects Allied to Medicine	91	103
C Biological Sciences	84	108
F Physical Sciences	85	129
G Mathematical & Comp Science	99	104
H Engineering	90	160
K Architecture, Building & Planning	97	105
L Social Studies	125	76
M Law	112	93
P Mass Communcation and Documentation	166	56
Q Linguistics, Classics & related	123	90
R European Languages, Literature & related	131	87
V Historical & Philosophical studies	85	114
W Creative Arts & Design	161	69
Z Combined Sciences	174	0
Y Combined Arts	98	101
Y Social Sciences Comb. Art	78	117
Z General	96	103
Y Sciences Combined Social Sciences	131	71

Table 5.14: University College London Index Scores Ethnic Profile of Courses (2004 Acceptances)

	White English	White Irish	White Scottish	White Welsh	Other white	Black Caribbean	Black African	Black Other	Asian Indian	Asian Pakistani	Asian Bangladeshi	Chinese	Asian Other	Mix W BC	Mix W BA	Mixed WA	Mixed O	Other	Unknown
UK Profile	85	40	11	36	326	67	113	104	300	134	277	395	504	127	201	245	267	331	45
Competitor Profile	78	76	11	38	301	213	263	287	327	227	490	271	418	197	245	156	244	413	123
Z Combined Sciences	108	0	0	0	632	0	0	0	200	0	0	0	304	0	0	0	0	0	0
Y Combined Arts	94	110	0	0	293	0	736	0	0	344	1031	573	0	0	0	152	412	606	0
YSocial Sciences Comb. Art	60	0	0	65	511	0	0	0	805	0	844	695	738	0	0	0	0	0	290
Z General	83	0	0	0	45	0	222	2443	518	349	0	407	244	0	0	326	0	305	56
Y Sciences Combined Social Sciences	43	0	0	0	199	0	0	0	644	598	2789	866	881	0	0	0	0	0	0
Y Combined social Sciences	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A Medicine & Dentistry	72	72	0	65	213	0	220	0	148	89	290	190	202	142	102	67	221	290	100
B Subjects Allied to Medicine	64	0	10	17	241	165	139	331	273	150	597	120	395	154	308	119	250	421	77
C Biological Sciences	69	120	11	13	282	591	360	0	349	239	709	476	338	114	492	295	201	269	152
D Vet Sci,Ag & related	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
F Physical Sciences	69	68	27	34	365	666	333	0	344	305	413	268	666	333	499	120	242	285	118
G Mathematical & Comp Sci	42	75	13	38	322	0	214	599	340	257	408	117	437	599	0	64	440	528	119
H Engineering	36	29	0	32	177	410	357	492	429	406	575	359	509	308	274	46	149	328	72
J Technologies	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
K Architecture,Build & Plan	76	70	16	0	274	189	172	0	267	0	631	81	526	0	189	218	158	789	76
L Social Studies	64	59	33	32	310	0	173	0	405	428	274	300	385	172	0	156	184	224	82
M Law	94	104	7	49	268	0	56	0	267	50	0	80	260	716	0	159	327	268	130
N Business & Admin studies	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
P Mass Comms and Documentation	33	0	74	0	619	0	0	0	650	0	1857	371	1857	0	0	0	265	0	232
Q Linguistics, Classics & related	94	171	0	70	154	168	0	727	338	0	218	0	291	128	467	126	112	682	111
R European Langs, Lit & related	78	60	13	49	233	0	298	0	191	318	0	477	191	136	0	254	393	294	143
T Non-European Langs and related	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
V Hist & Philosophical studies	84	71	19	44	270	0	386	858	615	351	468	143	542	112	0	97	64	245	185
W Creative Arts & Design	91	0	0	0	345	0	0	0	0	0	0	0	1743	0	0	0	381	1220	210

Table 5.15: University College London Index Scores and Percentage of Acceptances by Socio-Economic Group and Course (2004 Acceptances)[41]

		1	2	3	4	5	6	7
% Scores	% Acceptance UCL	32.02	27.15	12.4	5.13	2.05	6.62	2.13
	% Acceptance Competitors	30.36	30.51	12.1	5.06	2.96	7.11	2.52
	% Acceptance Nationally	18.03	25.57	12.32	6.01	3.96	10.73	4.59
INDEX SCORES	UCL compared with UK Profile	178	106	101	85	52	62	46
	UCL compared with Competitor Profile	105	89	102	102	69	93	84
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	A Medicine & Dentistry	174	107	82	45	33	47	11
	B Subjects Allied to Medicine	81	94	67	124	66	106	105
	C Biological Sciences	66	106	139	38	76	163	243
	F Physical Sciences	95	87	128	146	66	78	64
	G Mathematical & Comp Sci	103	77	124	127	36	115	109
	H Engineering	80	73	97	116	142	143	159
	K Architecture, Build & Plan	108	69	54	136	153	122	174
	L Social Studies	130	89	115	64	24	69	93
	M Law	141	72	91	106	0	52	146
	P Mass Comms and Documentation	57	131	0	104	0	309	0
	Q Linguistics, Classics & related	124	95	79	68	90	80	87
	R European Langs, Lit & related	119	84	98	110	76	82	88
	V Hist & Philosophical studies	106	85	196	174	80	59	27
	W Creative Arts & Design	81	83	107	0	295	152	197
	Z Combined Sciences	176	87	102	0	256	0	0
	Y Combined Arts	106	93	94	140	84	101	0
	Y Social Sciences Comb. Art	99	75	94	228	300	127	0
	Z General	102	101	157	136	0	88	0
	Y Sciences Combined Social Sciences	56	85	83	0	0	75	0

[41] Where 1 = Higher managerial and professional occupations, 2 = Lower managerial and professional occupations, 3 = Intermediate occupations, 4 = Small employers and own account workers, 5 = Lower supervisory and technical occupations, 6 = Semi-routine occupations, 7 = Routine occupations.

Table 5.16: University College London 2004 Index Scores and Percentage of Acceptances by POLAR Group and Course

		<16%	16% to 24%	24% to 32%	32% to 43%	>43%
% Scores	% Acceptance UCL	3.35	11.37	14.83	23.42	45.1
% Scores	% Acceptance Competitors	5.54	10.84	14.34	22.15	36.38
% Scores	% Acceptance Nationally	9.44	15.67	16.68	20.75	24.56
INDEX SCORES	UCL Compared with UK Profile	35	73	89	113	184
INDEX SCORES	UCL Compared with Competitor Profile	60	105	103	106	124
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	A Medicine & Dentistry	77	90	85	111	125
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	B Subjects Allied to Medicine	38	121	76	125	121
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	C Biological Sciences	98	106	151	93	113
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	F Physical Sciences	51	145	82	110	114
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	G Mathematical & Comp Sci	10	105	95	111	131
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	H Engineering	69	119	122	132	113
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	K Architecture ,Build & Plan	112	89	144	103	102
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	L Social Studies	21	119	84	83	127
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	M Law	68	91	107	103	152
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	P Mass Comms and Documentation	64	40	70	125	141
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Q Linguistics, Classics & related	107	85	111	89	120
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	R European Langs, Lit & related	52	110	121	81	119
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	V Hist & Philosophical studies	60	85	91	109	121
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	W Creative Arts & Design	68	99	103	107	111
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Z Combined Sciences	0	0	0	128	197
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Y Combined Arts	46	95	126	129	113
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Y Social Sciences Comb. Art	175	84	35	105	150
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Z General	26	71	90	121	131
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	Y Sciences Combined Social Sciences	135	104	166	80	148

Table 5.17: University College London Index Scores and Percentage of Acceptances by School Group and Course

		Comp. School	FE / HE	Grammar	Independent	Other	Six. Centre	Six. College	All State Schools
% Scores	% Acceptance UCL (less unknown)	20.39	11.48	11.73	42.48	1.75	1.12	11.05	57.52
	% Acceptance Competitors (less unknown)	34.93	9.96	9.98	29.36	1.21	0.89	13.67	70.64
	% Acceptance Nationally (less unknown)	33.6	30.85	7.06	12.38	1.63	1.02	13.47	87.62
INDEX SCORES	UCL Compared to UK Profile	62	38	171	354	111	113	84	68
	UCL Compared to Competitor Profile	54	107	109	134	134	117	75	76
UCL SUBJECTS COMPARED TO COMPETITOR INSTITUTIONS	A Medicine & Dentistry	64	78	109	127	263	123	62	81
	B Subjects Allied to Medicine	50	92	126	110	136	129	81	76
	C Biological Sciences	54	137	96	127	132	311	72	79
	F Physical Sciences	52	180	90	101	111	105	87	80
	G Mathematical & Comp Sci	58	106	72	129	133	150	97	77
	H Engineering	77	126	106	99	279	168	79	92
	K Architecture, Build & Plan	52	124	78	146	95	189	63	70
	L Social Studies	55	72	98	148	66	118	74	69
	M Law	52	112	111	208	0	0	44	65
	P Mass Comms and Documentation	0	151	143	253	0	0	30	43
	Q Linguistics, Classics & related	41	95	123	144	192	66	60	69
	R European Langs, Lit & related	64	100	90	137	0	68	87	76
	V Hist & Philosophical studies	54	65	166	125	110	0	75	76
	W Creative Arts & Design	12	344	53	28	0	0	0	117
	Z Combined Sciences	33	145	329	130	1174	0	114	113
	Y Combined Arts	71	30	61	136	0	0	100	68
	Y Social Sciences Comb. Art	39	36	87	187	328	0	79	59
	Z General	55	58	176	94	0	0	144	91
	Y Sciences Combined Social Sciences	76	65	65	177	697	0	97	84

Profiling acceptances by Mosaic highlighted which neighbourhood groups were under or over represented at University College London when compared to its main competitor institutions. Because the index scores are compared within JACS subject groupings, these indexes account (hold constant) the propensity for particular neighbourhood groups to study certain subjects (See Chapter 4). Thus, the index scores are to some extent a function of the geography of those neighbourhoods close to University College London, but also indicate how well University College London does at recruiting from these particular groups. It should also be noted that institutional course mix varies across competitor institutions, and when comparing within subject grouping these aggregations will not be homogenous: thus institution *a* may have very different course mix compared to institution *b* despite these being in the same JACS group. These profiles are summarised in Table 5.18.

Table 5.18: University College London 2004 Mosaic Profile of Courses[42]

	A	B	C	D	E	F	G	H	I	J	K
UCL Compared to UK Profile	173	42	104	66	193	104	14	38	48	100	77
UCL Compared to Competitor Profile	105	51	104	105	195	221	29	68	66	94	62
A Medicine & Dentistry	111	37	103	90	149	133	38	78	0	134	63
B Subjects Allied to Medicine	104	69	91	92	208	194	28	99	113	116	25
C Biological Sciences	86	54	101	147	211	394	55	42	202	91	64
F Physical Sciences	110	44	101	128	287	450	23	75	107	46	37
G Mathematical & Comp Science	75	68	117	142	264	234	0	56	0	46	47
H Engineering	62	101	125	129	218	366	0	85	0	114	16
K Architecture, Building & Planning	119	28	75	120	204	270	118	115	86	58	63
L Social Studies	112	50	139	67	122	234	0	24	63	65	46
M Law	117	94	101	69	170	83	0	65	0	54	106
P Mass Communcation and Documentation	164	0	118	120	248	0	0	0	0	0	0
Q Linguistics, Classics & related	103	43	90	105	167	126	0	71	0	138	112
R European Languages, Literature & related	115	28	105	93	163	130	136	56	0	106	71
V Historical & Philosophical studies	127	36	82	76	186	127	0	71	0	101	83
W Creative Arts & Design	104	57	75	112	138	581	0	156	359	0	147
Z Combined Sciences	97	0	0	0	279	0	0	211	0	249	139
Y Combined Arts	92	23	98	90	174	0	0	118	0	141	173
Y Social Sciences Comb. Art	92	55	97	65	205	151	0	82	0	221	54
Z General	108	75	141	80	314	175	0	0	305	61	30
Y Sciences Combined Social Sciences	102	48	126	148	199	0	335	299	0	0	0

Figure 5.26 and Figure 5.27 illustrate the importance of comparing profiles within competitor groups. When comparing University College London with all institutions, University College London has a profile similar to those of most other top

[42] A = Symbols of Success, B = Happy Families, C = Suburban Comfort, D = Ties of Community, E = Urban Intelligence, F = Welfare Borderline, G = Municipal Dependency, H = Blue Collar Enterprise, I = Twilight Subsistence, J = Grey Perspectives, K = Rural Isolation.

universities, with an over representation of those neighbourhoods which are most affluent / advantaged. As shown in Chapter 4 these areas possess a propensity for higher achievement in those qualifications considered by University College London and its competitors when making offers. Furthermore, "Urban Intelligence" is also overrepresented, and might be attributable to the high occurrence of these neighbourhood groups within London.

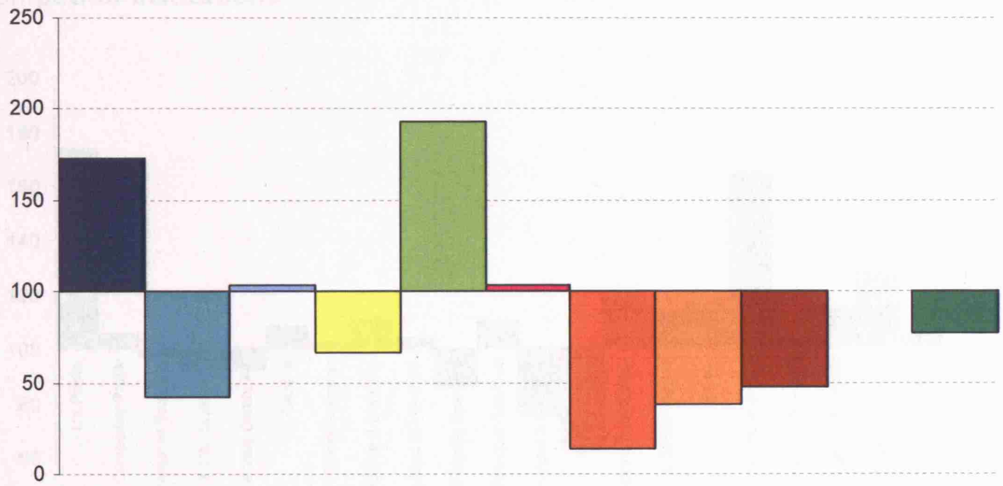


Figure 5.26: University College London Mosaic Index Score Profile Compared to all UK Institutions

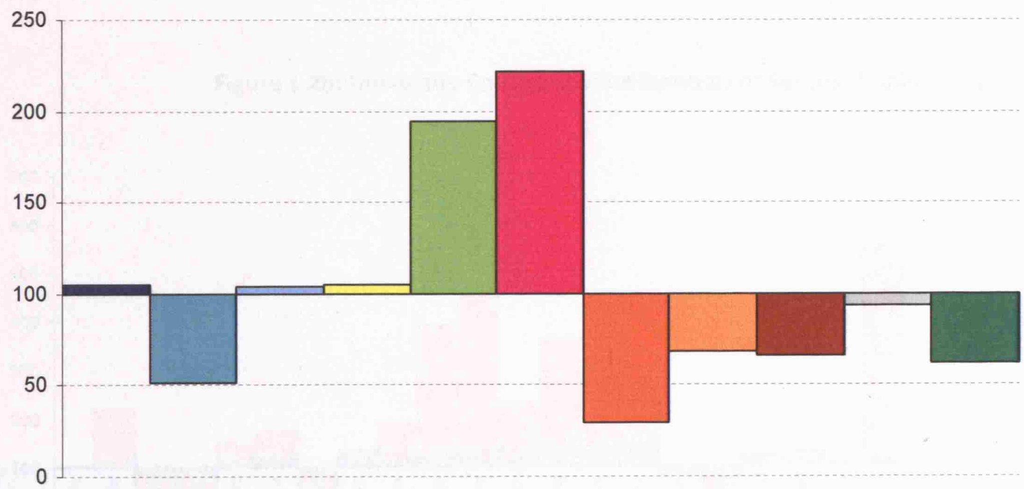


Figure 5.27: University College London Mosaic Index Score Compared to the Competitor Base

When comparing UCL with its competitor institutions as a base, UCL does not show significant over representation of "Symbols of Success" neighbourhood groups, despite high frequencies of these areas within London. UCL also recruits students from "Welfare Borderline" neighbourhoods twice as effectively (index ~ 200) as its competitor institutions. This is a significant achievement as these groups are underrepresented nationally. There are however large numbers of these areas to be found within London. Although the full results were shown in Table 5.18, graphs were created for each geodemographic Group. Figure 5.28 and Figure 5.29 show the propensity for students within particular neighbourhood Groups to accept places at University College London by JACS groupings when compared with competitor institutions.

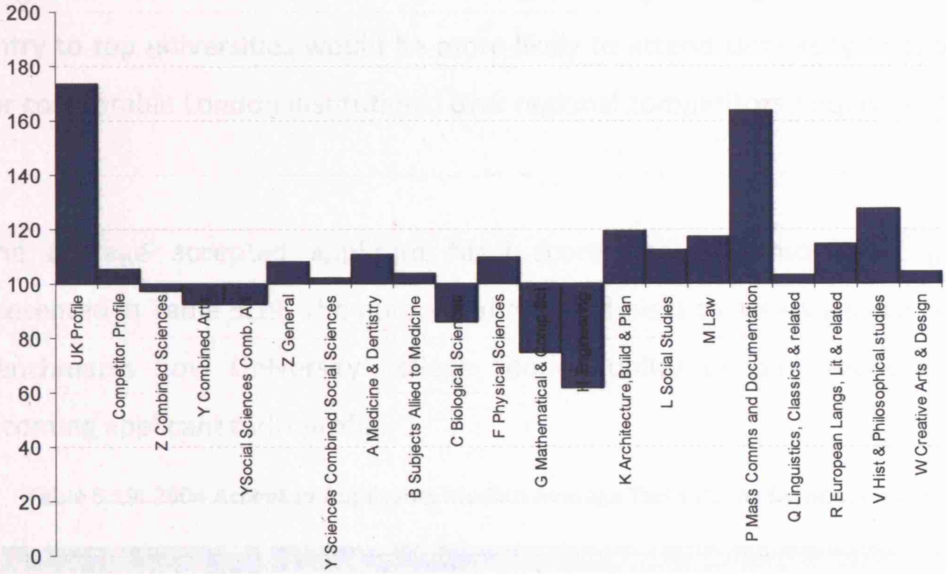


Figure 5.28: University College London Symbols of Success (2004 Acceptances)

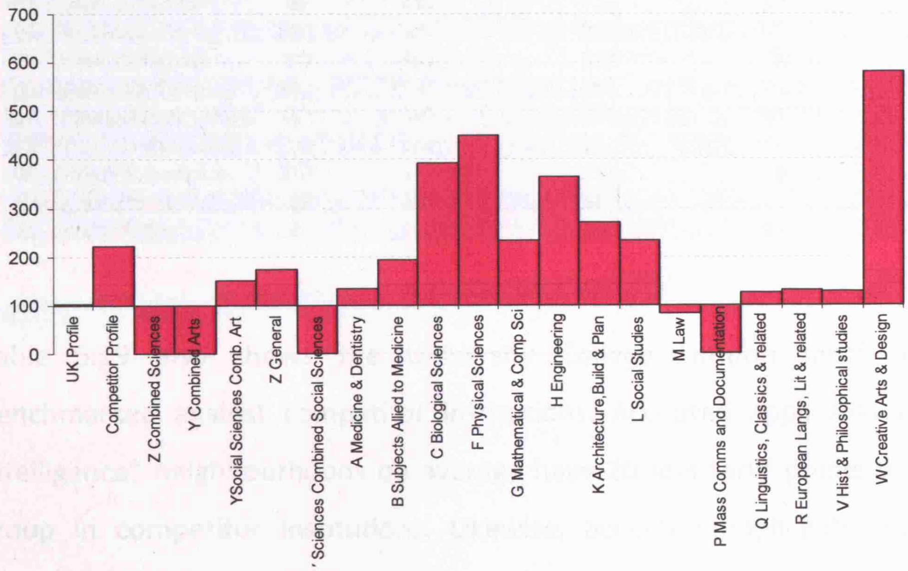


Figure 5.29: University College London Welfare Borderline (2004 Acceptances)

Two neighbourhood Groups that are over represented across almost all subject groups are "Welfare Borderline" and "Urban Intelligence". This is unsurprising as there is over representation of these neighbourhoods in close proximity to University College London, and it is these neighbourhoods who nationally travel shorter distances to accept offers (see Section 4.3.1). Furthermore, nationally, students from within "Welfare Borderline" neighbourhoods have a low propensity to travel long distances to accept places, and as such those from within these neighbourhoods in London who achieve grades high enough to be considered for entry to top universities would be more likely to attend University College London (or comparable London institutions) over regional competitors / equivalents.

The average accepted applicant tariff scores by geodemographic group are presented in Table 5.19. The purpose of this analysis is to assess against competitor benchmarks how University College London policy on admissions affects the incoming applicant tariff profile.

Table 5.19: 2004 Accepted Applicants Median Average Tariff Points by Geodemographic Group

	Average Tariff Points				
	All	**Competitor**	**UCL**	**Competitor. Difference**	**Frequency of UCL Accept. in 04**
01 - Symbols of Success	330	420	420	0	878
02 - Happy Families	260	410	430	20	132
03 - Suburban Comfort	280	410	420	10	542
04 - Ties of Community	200	390	390	0	246
05 - Urban Intelligence	200	390	370	-20	286
06 - Welfare Borderline	80	330	360	30	111
07 - Municipal Dependency	150	340	420	80	11
08 - Blue Collar Enterprise	200	384	385	1	68
09 - Twilight Subsistence	210	389	335	-54	14
10 - Grey Perspectives	280	410	410	0	135
11 - Rural Isolation	310	410	430	20	148

Table 5.19 also shows the University College London tariff score profile benchmarked against competitor institutions. Accepted applicants from "Urban Intelligence" neighbourhoods on average have 20 less tariff points than the same group in competitor institutions. Likewise, accepted applicants from "Welfare Borderline" neighbourhoods have on average 30 more tariff points compared with competitor institutions. Acceptances from "Municipal Dependency"

neighbourhoods have average scores of 80 more tariff points at University College London than those of the same group in competitor institutions. These scores are based on a very small number as these applicants predominantly come from deprived neighbourhoods outside of London. These analyses can also be conducted within course groupings (See Table 5.20). Accepted applicants from "Welfare Borderline" neighbourhoods have already been shown to be overrepresented in Engineering, Physical Sciences and Biological Sciences compared with University College London competitor institutions (See Table 5.18). It can also be seen that the lower accepted applicant tariff profile (all negative) partially explains these higher index scores.

Table 5.20: 2004 Accepted Applicants Median Tariff Points by Preferred Course[43] Type

	Average Tariff Points				
	All	Competitor	UCL	Competitor Differences	Frequency of UCL Preferred Course Acceptances in 04
Z Combined Sciences	270	380	435	55	5
Y Combined Arts	300	410	400	-10	46
Y Social Sciences Comb. Art	290	410	340	-70	37
Z General	220	480	420	-60	12
Y Sciences Combined Social Sciences	220	407.5	440	32.5	20
A Medicine & Dentistry	460	460	460	0	401
B Subjects Allied to Medicine	240	350	340	-10	134
C Biological Sciences	280	400	390	-10	191
F Physical Sciences	320	400	370	-30	193
G Mathematical & Comp Sci	200	440	420	-20	118
H Engineering	260	400	360	-40	137
K Architecture, Build & Plan	240	400	380	-20	103
L Social Studies	266	400	450	50	285
M Law	310	460	480	20	133
P Mass Comms and Documentation	240	370	330	-40	4
Q Linguistics, Classics & related	350	420	420	0	175
R European Langs, Lit & related	360	390	400	10	126
V Hist & Philosophical studies	340	410	420	10	240
W Creative Arts & Design	200	370	340	-30	39

The interactions between course and neighbourhood type were further explored in Table 5.21 through Table 5.24. In Table 5.23 counts of University College London acceptances within the cross tabulation are presented. Where the acceptance count was less than or equal to 3 in Table 5.23, those corresponding difference

[43] Derived from a general preference in course group based on all applications from an accepted applicant.

University College London could identify areas within the college where further support might be required to extend access to underrepresented groups. Table 5.6 has fields with 0 index scores which indicate where a department recruited no students from these neighbourhood Groups. Care should be taken when interpreting departmental index scores, as the low spread of accepted applicants across numerous University College London departments can cause volatility in the scores. An example of these effects can be illustrated by the over representation of the "Rural Isolation" group in the Dutch Department. There are only 3 total acceptances recorded to this department and therefore the level of reliability of the index scores will be limited.

Table 5.25 shows that the Faculty of Social and Historical Sciences has an over representation of "Symbols of Success", and an under representation of "Welfare Borderline". There are no "Municipal Dependency" acceptances recorded. In isolation: these data may be interpreted as a faculty requiring further assistance to widen access. However, when these data are compared to national profiles for students to study these courses, the over representation is less surprising considering that the types of courses within this Faculty include those from JACS groups F, L, and V, all of which have a national trend for overrepresentation of the "Symbols of Success" neighbourhood Group. However, some people would argue that all courses of Higher Education should appeal (or be seen to be available) equally to all groups within society. These inequalities could have multiple causes including promotion by the subject area, or applicant prior educational attainment limiting chances of application success to these subjects. Key "widening participation" applicants as defined by HEFCE would be those whom are most likely to be in the POLAR group <16% participation, not gone to an independent school and are in the lowest socio-economic group. Chapter 4 demonstrated that applicants from "Welfare Borderline" neighbourhoods fall within these categories most prevalently and as such can be used to rank departmental contribution towards recruiting "widening participation" applicants (See Table 5.27).

5.6 Conclusion

This chapter has introduced Educational Market Profiler which forms a prototype decision support service which could be deployed by UCAS to provide centralised and freely available information to Higher Education institutions. It presented a review of those data and services currently available, however concludes that these do not meet the information requirements of an increasingly market orientated sector. The Educational Market Profile was grounded in theory and empirical observations presented in previous chapters, and designed in software which integrates with UCAS current data infrastructure. This decision support tool was presented thematically by function including geographical analysis, schools analysis and institutional benchmarking, with a case study in each section for University College London. This tool includes a range of conventional indicators which were shown to be either influential or a direct a measure of stratified access. Furthermore, geodemographic profiles were constructed using Mosaic as UCAS possessed a licence for its use, however this could be undertaken using other commercial classifications. The major unique contribution demonstrated by Educational Market Profiler is the integration of data from the Higher Education sector and schools, enabling profiles to be built for both those who attend, and those who do not attend Higher Education.

PART 3: DESIGNING GEODEMOGRAPHICS FOR HIGHER EDUCATION

6

CREATING OPEN SOURCE GEODEMOGRAPHICS - REFINING A NATIONAL CLASSIFICATION OF CENSUS OUTPUT AREAS FOR HIGHER EDUCATION

6.1 The Public Sector and Geodemographic Classifications

There are a variety of commercial geodemographic classifications which can be used to inform spatial decision making in Higher Education, however none have been specifically designed for this purpose. Chapter 3 described how geodemographic classifications originated in the public sector (as a method of targeting deprived areas (see Webber 1977; Webber 1978; and Webber and Craig, 1978) however were subsequently augmented with private sector data for commercial applications such as customer segmentation and direct mailing (Harris *et al*, 2005). Public sector data should be made more readily available, however even with its current limited dissemination no current commercial vendor incorporate public sector data at clustering unit level into commercial classifications. Experian were the first commercial vendor to provide a public sector version of their commercial classification, however this has only been attained superficially at the level of "pen portraits" which provide additional descriptive material of the clusters. Although these classifications may be promoted as bespoke solutions for the public sector, they do not address a number of key concerns including whether it is appropriate for a general purpose classifications describing private consumption of the goods and services supplied by the private sector should

In socio-economic clustering applications such as geodemographics, it is proposed that dissection is a more accurate representation of the function of the clustering algorithm. The input data for the cluster analysis has high dimensionality across multiple attributes, and the range of values associated with these variables can vary widely across different spatial locations. On this basis, the logic of a fuzzy classification (dissection) is reinforced, as different discrete dissections of the data could result in multiple classification schemes. With a fuzzy classification the blurring of boundaries between clusters is accommodated through a membership statistic which shows how well the data points fit within particular groupings. Although a discrete classification "can be a gross over simplification of the structure in the data set" (Gordon, 1981:58), the prevalence of these classification types in commercial geodemographic applications has remained unanimous. An explanation for this could be that end user communication is easier for discrete rather than for fuzzy classifications. For example, it is easier to describe that someone resides in neighbourhood groups *a* or *b*, rather than a mix of multiple groups, and can be particularly confusing when ambiguous assignments are supported by supplementary descriptive materials. The Output Area Classification (OAC) comprises a set of discrete hierarchical groupings; however, in recognition of the advantages of fuzzy clustering methods, OAC is supplemented with membership statistics that may be interpreted as a fuzzy geodemographic classification.

Before running a cluster analysis data must be standardised to reduce the effect of outliers and measure the data on the same scale, in order to ensure that all variables have the same weighting. Romesburg (1984: 78) discusses how standardisation prior to clustering prevents the units used for attribute measurement from affecting the similarities between objects, and therefore allowing more equal contribution by each of the variables. There are many ways in which the data could be standardised and these are evaluated by Vickers *et al* (2005), who found that in the production of OAC, range standardisation method (Wallace *et al*, 1996) performed most effectively at reducing outlier effects in Census based Output Area (OA) level classification. This method standardises all the data between 1 and 0. The OAC classification is built using a *k*-means algorithm which partitions an *R* multidimensional data matrix into *k* clusters or groups based on local optimisation criteria. For a full evaluation of why this algorithm is most suited to clustering applications at Output Area geography see Vickers (2005).

The *k*-means method (MacQueen, 1967) is an iterative relocation algorithm, which assigns the OA data points into *k* clusters based on a standardised Euclidean minimum distance metric. The algorithm sets the initial location of the cluster centroids as the random *k* OA data points. The distance of the data points to each cluster centroid is then calculated, and each data point is provisionally assigned to its nearest cluster centre. A clustering criterion statistic is then applied to measure the homogeneity within these temporary cluster allocations. Pythagoras' Theorem is applied in a *Rn* dimensional space (Gordon, 1981) where a dataset vector comprises of *n* variables (dimensions) weighted by an associated population size. At each iteration of the model the population weighted distance between the data points and the cluster centroids are re-calculated using Equation (6.1).

$$d_{iQ} = \sum_{p=1}^{n} (x_{iR} - x_{QR})^2$$

(6.1)

This equation is applied to all clusters within a single iteration of the model, i.e. *kn* clusters, and it is the sum of these results which form *f* Equation (6.2), a model objective function assessing the overall model performance or within cluster homogeneity.

$$f = \sum_{k=1}^{n} (d_{iQ})_k$$

(6.2)

After the first iteration of the model where initial cluster centroids (seeds) are randomly placed and all data points are temporarily assigned to their nearest seed, the *k*-means algorithm attempts to find a local optimum through an objective function that reallocates data points iteratively from their initial assignments. Each

2. Use the existing OAC classification to create a further tier in the hierarchy by re-clustering each Sub Group into a number of "Micro Groups", and then by adding sector data re-cluster these into a bespoke educational classification.

There are a number of problems associated with the former option. Firstly, the classification would be re-created from base principles and therefore have no likely resemblance to the existing Output Area Classification. This could result in difficulty when communicating these groups to end users of the existing classification. Secondly, unlike the variables used in the OAC classification taken from decennial census data, education data have a smaller coverage socio-economically and an uneven geographical coverage, resulting in the increased prevalence of outlier values based on small base counts. These effects could be reduced though standardisation and careful population weighting, but they would still likely cause undesirable noise in the resulting classification. The second option potentially resolves these issues by providing a link to the existing classification and also a method by which small base count effects may be minimised because the clustering input are segments, i.e. aggregations of OAs containing similar socio-demographic characteristics Therefore, the first stage in the analysis was to create a new tier in the OAC classification. A similar product called Mosaic Segments (Experian, Nottingham, UK) provides a 243 cluster solution which is a finer disaggregation of its 61 Mosaic Types. Previous work using this classification (Singleton and Farr, 2004) has demonstrated that this level of aggregation is effective for re-clustering of education data.

The input data used to create this new finer level classification consisted of a series of standardised 2001 Census variables at Output Area level, and was the same data used to construct OAC. This UK dataset was split into 52 separate groupings of Output Areas based on their assignments in the OAC subgroup classification. These 52 datasets were separately re-clustered using the *k*-means algorithm implemented in the SAS statistical software[46]. SAS offers more powerful data manipulation functions and scripting than other leading statistical software, and also is the database format used by the Higher Education Statistics Agency (HESA) – and thus its use reduces the possibility of data corruption through converting the database into alternate formats.

As discussed earlier in this chapter, a method to derive a globally optimised local model for a single dataset is to run the *k*-means cluster analysis to convergence with multiple initial random seed locations, each time comparing the performance of the final classification using optimisation criteria such as the *R*-squared statistic. The flow chart in Figure 6.5 pictorially represents this process and the application to perform the analysis is written in the SAS code presented in the Appendix Figure 12.12. In order to avoid the problem of poor initial seed selection discussed earlier, each of the 52 datasets had *k*-means algorithm repeat run 10,000 times with total processing time taking roughly 144 hours to complete on a workstation with 2 gB RAM and a 3.2 gB Pentium 4 processor.

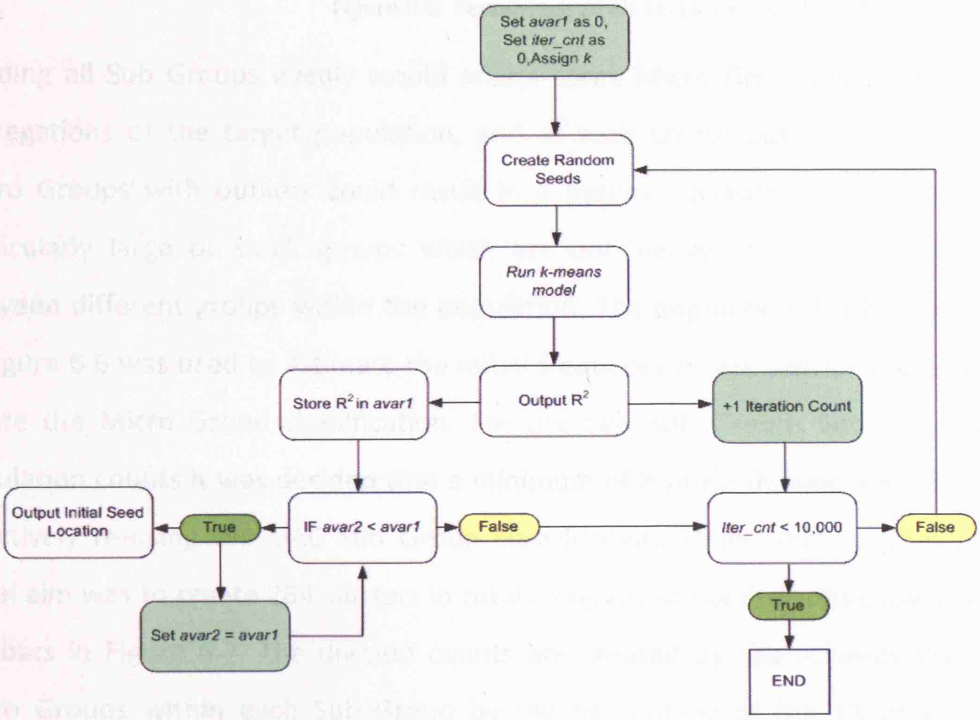


Figure 6.5: Flow Chart of the Clustering Optimisation Process

[46] SAS is a commercially available statistical software: http:www.sas.com

Following the methodology by which the hierarchies of OAC were created, each of the 52 Sub Groups were divided into a further level referred to as "Micro Groups". The aim was to create a classification with a similar frequency of clusters to commercial "segment level" classifications. The initial frequency of divisions was based on the distribution of 18-19 year olds within the OAC Sub Groups. The percentage distribution of 18-19 year olds by OAC Sub Groups can be seen in Figure 6.6.

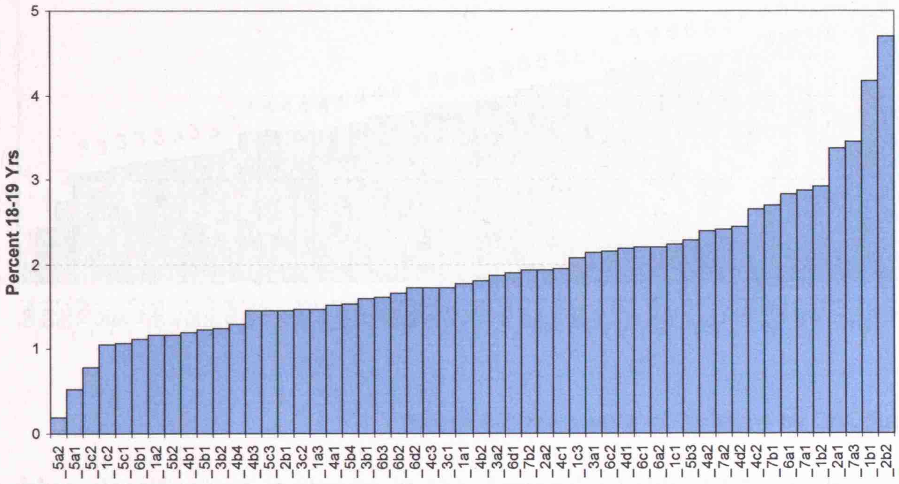


Figure 6.6: Percentage of all 18-19 Year Olds by OAC Sub Groups

Dividing all Sub Groups evenly would create some Micro Groups with very small aggregations of the target population, and as such create outliers. Re-clustering Micro Groups with outliers could result in a bespoke classification consisting of particularly large or small groups which are not ideally placed to discriminate between different groups within the population. The population distribution chart in Figure 6.6 was used to estimate the initial frequency of the divisions required to create the Micro Group classification. For the two Sub Groups with the lowest population counts it was decided that a minimum of a single division was required, effectively re-using the OAC Sub Group classification. Under this constraint, the initial aim was to create 264 clusters in total using the initial divisions shown above the bars in Figure 6.7. The division counts are created by apportioning the total Micro Groups within each Sub Group by the percentage of the 18-19 year old population present. The actual figures are represented using the green bars, and

$$z = \frac{x - \bar{x}}{\sigma} \tag{6.3}$$

This equation divides the difference between a case variable (x) from the mean of the total cases in the population ($\bar{x}$) by the standard deviation of all case values (σ).

The choice of standardising with z-scores over other methods (such as range standardisation and log scores, used in the original OAC classification) is to prevent the scores being capped, and as such allow possible outliers to influence the final classification as these may demonstrate important Higher Education behaviours. Because the units of clustering are larger than in OAC, i.e. Micro Groups rather than OAs, outliers created by aggregating data by small geographical areas are greatly reduced, and as such any outliers present in the data matrix should represent significant features which should be picked up in the final cluster assignments. The process of filtering small aggregation outliers from geographical features is helped in the clustering by weighting each case (Micro Group) by the total population for these areas, therefore reducing the weight / influence of those areas based on a smaller sample size.

An Higher Education specific geodemographic classification could feasibly serve a number of different purposes including marketing, extending access, widening participation or subject specific targeting. When selecting input variables these purposes were considered with the aim of creating a classification useful for a variety of tasks demanded by Higher Education decision makers. Variables selected are chosen to provide both indirect and indirect indicators (see Section 3.3) relating to those constructs which influence participation in Higher Education. The aggregate measure is likely to be suitable for profiling both access to aggregate Higher Education and disaggregated to predict course and institutional profiles. These applications are tested in Chapter 7.

6.3.2 Students and Distance

The use of distance travelled to accept a degree place is a proxy for geographic restrictions which are apparent in applicants from lower socio-economic groups, either through the financial cost of travel to a disparate institution or the social networks which bind them to their local community. Distance is measured in this analysis using a straight linear path between the accepting institution and the student's home; however other distance metrics could be used. Travel may for example be more prevalent in areas which demonstrate increased accessibility to transport networks. In this analysis a straight linear distance was used as a pragmatic way of gauging relative variability between neighbourhoods, without more complex analysis involving transportation data. The co-ordinates for these locations are derived from the All Fields Postcode Directory for 2001. Once two pairs of *xy* co-ordinates are derived, where *i* is the student home and who is attending *j* institution, Pythagoras Theorem was used to calculate the distance using Equation (6.4).

$$d = \frac{\sqrt{(x_i - x_j)^2 * (y_i - y_j)^2}}{1600}$$

(6.4)

Including distance in the classification helps discriminate between those who may not participate because they are remotely located from institutions and those who may not participate because of their socio-economic profile.

6.3.3 Average A-Level Scores & Non A-levels

In the 2001 HESA data A-levels were measured on a points scale, ranging from 10 points for an A grade to 2 points for an E. These scores were cumulative, so someone attaining AAE would score 10+10+2 = 22 points. These scores were aggregated by each OAC Micro-Group to create mean prior attainment scores. Prior attainment, particularly with regard to academic qualifications such as A-Level have been seen as "key to the reaffirmation of middle class privilege in education and employment" (Leathwood and Hutchings, 2003:153) and as such should provide a good discriminator of neighbourhood disadvantage. Where these scores are not recorded in the HESA data, the applicant will have qualified for Higher Education through a non A-Level qualification; these have been recorded separately in this analysis as a non-A level variable.

6.3.4 Social Class

In 2001 HESA data social class was measured on the Registrar General's Social Scale which groups occupation into 7 different categories. These categories are:

- I Professional occupations
- II Managerial and technical occupations
- IIIN Skilled non-manual occupations
- IIIM Skilled manual occupations
- IV Partly-skilled occupations
- V Unskilled occupations
- VI Armed forces

Low rates of participation by the lower social classes have been documented ever since the Robbins Report (Robbins, 1963), and the extent to which these social barriers have been successfully addressed is debatable. In order for the classification to discriminate between the higher and lower social class groups, a variable was created from frequency of students within the groups IIIM, IV and V.

6.3.5 Degree Subject Chosen

The 2001 HESA data use the Standard Classification of Academic Subjects (SCAS) to aggregate individual courses into subject groupings. The extent to which different neighbourhood types participate across these subjects is essential for both marketing and widening participation. The inclusion of the proportion of students with which it is possible to both describe and discriminate between groups within society. The educational classification will therefore be created with each Micro-Group having a population weighting assigned.

Before performing a cluster analysis the data were explored to examine the correlations between the variables. It can be argued that highly correlated variables within a cluster analysis results in data redundancy and can have undesirable effects in final cluster assignment (Vickers, 2005). Harris *et al* (2005) also discuss how it is important that included variables add new information rather than repeating what is already known. It is claimed that the methodology employed by Experian in the construction of Mosaic allows correlations to be maintained, through the use of weighting variables where necessary in order to improve the classification. Correlation between variables has the effect of adding extra weight to a particular dimension of the classification, and depending on which group of variables are being measured could actually benefit the final assignment of clusters. For example, a key purpose of the Higher Education classification under development here is discrimination between areas that have high and low participation. One would expect the various factors which lead to these patterns of inequality to be highly correlated, for example A-Level Points and Social Class. Both of these dimensions contribute towards low participation and as such should be included. Their potential correlation re-enforces an important dimension of the classification and as such should be allowed to manifest itself in the final cluster assignment. In a classification where a broad range of variables is included to inform a number of roles, the correlation between small subsections of these data reduces the need for variable weighting. The converse argument is that these dimensions could be artificially created through the use of effective weighting. The main issue with variable weighting to inform a classification is that these choices are inherently influenced by those priori models held by the researcher of how the final classification will represent the data. If a classification builder wished a particular dimension to be over represented, for example the existence of wealthy Asians versus poorer Asians, this could quite easily be achieved through variable weighting.

The Output Area Classification (OAC) minimised the inclusion of correlated variables and did not use weights. The bespoke educational classification developed here only includes relevant variables, some of which are correlated, and like OAC did not use variable weighting for the reasons discussed above.

In order to understand the correlation within the input dataset, a matrix of all input variables was created using a population weighted Pearson Correlation Coefficient (See Table 6.8). Cells with a red arrow icon indicate a negative correlation greater than 0.6, a green arrow indicates a positive correlation greater than 0.6 and all other values have a yellow arrow. Highly positively correlated variables include A Level Points, independent schools, distance travelled to accept a place and young participation. As one would expect, each of these variables is highly negatively correlated to low Social Class and no A-levels. These patterns are unsurprising and refer to a core component of the classification, to discriminate between those areas of high and low participation. One would also expect these variables to correlate with some subjects as entry grades vary between subject groups, subjects appeal to different people and subjects are not evenly distributed across Higher Education institutions. For example there is a high negative correlation between low Social Class and participation in JACS Group A Medicine and Dentistry.

6.5 How Many Clusters Should an Educational Classification have?

The k-means algorithm clusters the input data matrix into the k groups that are specified by the researcher. Therefore, unless a prior model of how many groups should exist within the dataset is known, a method of selecting a sensible cluster frequency is required. As mentioned earlier in this chapter, one method of doing this has been demonstrated by Debenham (2001), and entails running the k-means algorithm for multiple iterations of k and plotting the average distance between the data points and their closest cluster centroid. These charts show the homogeneity of each cluster solution against the number of clusters. The higher the number of clusters, the smaller the mean distances between the data points and the nearest cluster centroid. The charts thus illustrate the trade-off between mean distance and classification complexity. Debenham (2001) conducts this analysis by running only a single cluster analysis for each k value. This has the disadvantage described earlier that the k-means algorithm is sensitive to the location of initial seeds – a problem that can be largely circumvented through repeated analysis using multiple initial seed values. Debenham (2001) selects a final k value based on interpretation of apparent breakpoints in the plot of cluster homogeneity against number of clusters. However, without re-running the cluster analysis, these observations may be anomalies based on inappropriately selected initial random seeds. Although this method is useful in principle, it needs to be adapted in order to provide more robust results.

The method adopted in this study builds on Debenham (2001) and runs the algorithm for k_{n-2} models where n is the total number of Micro Groups within the dataset (176). However, in order to improve the confidence with which the trade off between cluster homogeneity and numbers of clusters is made, each iteration of k was re-run 10,000 times. The median, minimum and maximum distances and overall R-Squared were averaged over the 10,000 iterations for each k value and are graphed in Figure 6.9 and Figure 6.10. In these Figures, the dark line represents the median and the grey whiskers the minimum and maximum values. The systematic location of the median towards the maximum values in Figure 6.9 indicates that there are more iterations of the model close to the maximum performance than the minimum performance.

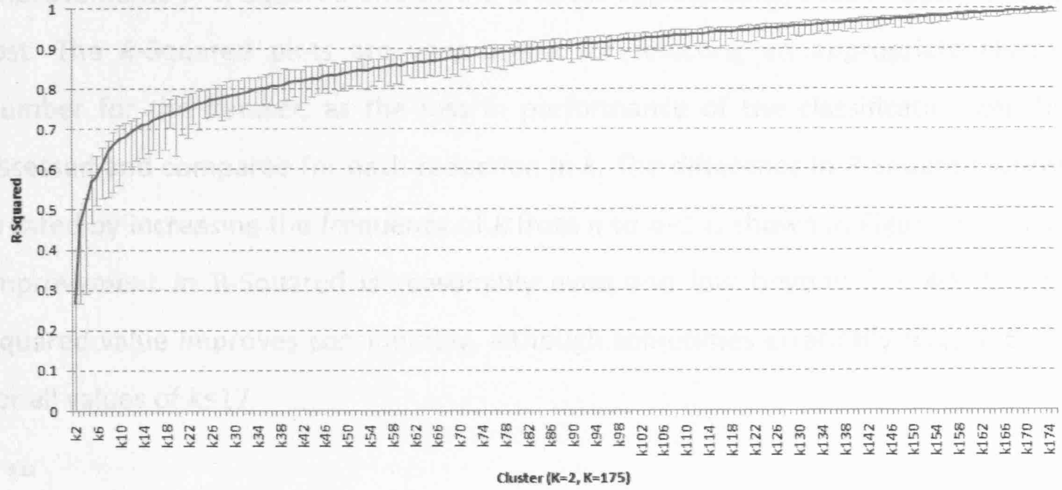


Figure 6.9: Cluster Performance Measured by R-Squared Scores (*k*= 2 – 175)

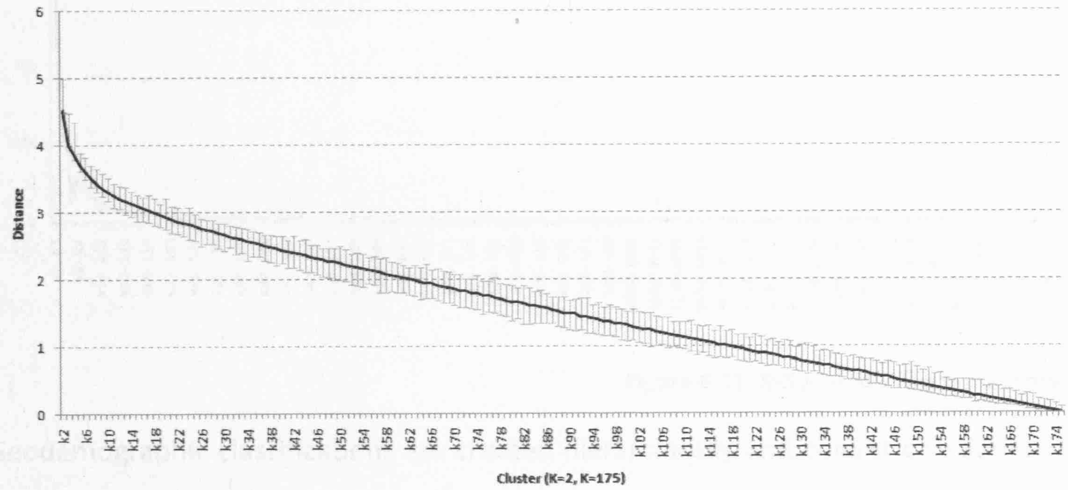


Figure 6.10: Cluster Performance Measured by Distance Scores (*k*= 2 – 175)

These graphs show that the *R*-Squared increases with the number of clusters specified, although not in a linear fashion. Furthermore, as *k* decreases so the variability of the *R*-Squared increases, providing further justification of the need for multiple model runs to attain robust information, particularly at lower values of *k*. The increased variability in *R*-Squared at most of the lower *k* values is caused by the grouping of the data points into smaller aggregations and this corresponds to a

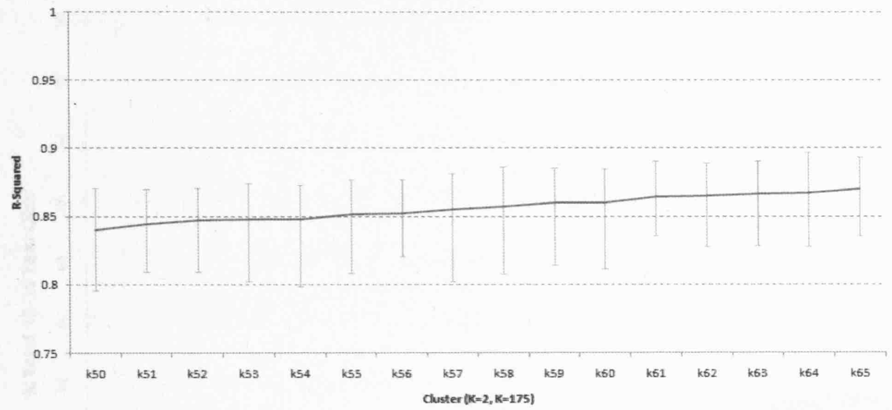


Figure 6.12: *R*-Squared Results from 10,000 Cluster Analysis

Each of these assignments of *k* appears to perform well at discriminating within the input data matrix and, as demonstrated in the earlier exploratory analysis, the minimum and maximum bars further illustrate the need to optimise each *k* allocation. The total and 18-19 year old population from the 2001 Census were then aggregated into the *k*=50 to *k*=65 models in order to ensure that no outliers of this key target population had been created in the clustering process. The requirements analysis suggests that a more even distribution of 18-19 year olds, the core applicant group, between clusters is better, all other things being equal. The model demonstrating the most even distribution of 18-19 year olds across the new clusters was *k=53* (see Figure 6.13) and as such was chosen as the final model. It should however be noted that the distribution of 18-19 year olds is still skewed. The red line is drawn at 1.89 which would divide the principal applicant group equally between the 53 clusters (i.e. 100/53). The uneven distribution of household and population counts is characteristic of most geodemographic classifications and in the Mosaic commercial classification the percentage assignment of total households to Mosaic Types ranges from 0.17% - 3.82%.

The variables for which these index scores were calculated included:

- Propensity for course level participation.
- Propensity to attend a Russell Group[48] institution
- Propensity for Group to appear in POLAR participation Wards
- Propensity for the OA Group to contain particular Mosaic Groups

These statistics provide useful material for end users to describe the general characteristics of students within each of the educational OAC neighbourhood Groups, and also a method of external validation through cross classification with Mosaic Groups.

6.8.1 Group A

Group A are the very lowest participation neighbourhoods, generally found within urban areas outside of London. Students who do attend university from these areas have a high propensity to study either Education or a Biological Science. Attendance at Russell Group institutions is half the national average.

6.8.2 Group B

Group B neighbourhoods are low participation, again found predominantly outside of London. Those students who do attend university also have a high propensity to study Subjects Allied to Medicine, Education or Creative Arts and Design. Few students in these areas attend Russell Group institutions.

[48] The Russell Group is as association of leading UK research-intensive Universities whose membership include: University of Birmingham, University of Bristol, University of Cambridge, Cardiff University, University of Edinburgh, University of Glasgow, Imperial College London, King's College London, University of Leeds, University of Liverpool, London School of Economics & Political Science, University of Manchester, Newcastle University, University of Nottingham, Queen's University Belfast, University of Oxford, University of Sheffield, University of Southampton, University College London, University of Warwick

6.8.3 Group C

Participation in Group C neighbourhoods is low, but markedly higher than Groups A and B. These neighbourhoods are predominantly found on the outskirts of larger urban areas in the South and Midlands of England. A high proportion of participating students study Mathematics and Computer Science. Over-all participation in Russell Group institutions is low.

6.8.4 Group D

Group D neighbourhoods are predominantly rural, with a moderately high participation rate. As one might expect of rural residents, many students study Veterinary Science or Agriculture. Other popular course choices include Engineering, Technologies, Architecture/Building/Planning and both European and Non European Languages. Students from this Group are 25% more likely to attend a Russell Group institution than the national average.

6.8.5 Group E

Group E neighbourhoods are found across all main urban areas (although only the periphery of London). They have average to low participation rates but Education is a popular subject of study. Other subjects that are reasonably popular include Biology, Physical Sciences, Mass Communications / Documentation and Creative Arts / Design. This Group has an average propensity to attend a Russell Group Institution.

6.8.6 Group F

Group F are frequent participants in Higher Education, and are found in both rural and urban areas across England, predominantly in affluent neighbourhoods. They exhibit no strong course preferences, but are 17% more likely than the national average to attend a Russell Group institution.

6.8.7 Group G

Group G populate affluent neighbourhoods in the suburbs of many large urban areas, particularly in London and the South East of England. Students from these areas are very likely to go to university, and have a tendency to study Medicine, Architecture, Building & Planning, Social Studies, Languages, Philosophy or History. Students from these neighbourhoods are 50% more likely to study at a Russell Group institution than the national average.

6.8.8 Group H

The predominantly rural neighbourhoods in Group H are the most likely of all Groups to participate in Higher Education, and students from these neighbourhoods are 66% more likely to study at a Russell Group institution than the national average. Subjects which are popular with students from these areas include Veterinary Science, Physical Sciences, Engineering, Social Studies, Business Studies, Languages, History and Philosophy.

6.8.9 Group I

Group I neighbourhoods have average to low participation rates, and mainly come from urban areas, especially Inner London. Students from these areas study Computer Science and Mathematics and a range of other subjects. These students are unlikely to attend a Russell Group Institution.

6.8.10 Group J

Group J neighbourhoods have average rates of participation. They are mainly found in large urban areas across England, especially in the suburbs of London. Many students study Medicine or Subjects Allied to Medicine, Mathematics, Computer Science, Law, and Business. Students from these areas are 30% less likely than the national average to attend a Russell Group Institution.

6.9 Conclusion

This chapter has demonstrated a pilot method by which bespoke classifications for a particular sector or application can be created using public sector data sources. The motivation for this analysis lies in the observation that typologies created by commercial classification providers supply no evidence to justify why the inclusion of data relating to private consumption of goods is appropriate for predicting public consumption. Furthermore, the exact nature of the weighting schemes and data used to derive such commercial classification systems is closed to the public, which is of concern in public sector applications which may apportion real life chances, rather than just material consumer offerings. The addition of Higher Education sector data is seen as a positive step beyond use of generic and re-labelled classification for purposes for which they were not originally designed, and a challenge to the implied assumption that that individual use of public services such as education should directly correspond with the ways in which consumers use private goods and services. This work also responds to concerns that the data inputs used to create generic commercial geodemographic classifications come from disparate private sector and closed sources, their provenance is often unknown, and that the assumptions used to create such classifications cannot be scrutinised or tested by end users. The negative potential social implications of using such classifications in areas of public service provision should not be under-estimated, and potentially significantly reduce the life chances of stakeholders in public services. The methodology has shown how a classification built using the 2001 Census can be refined for a specific purpose through the augmentation of sector specific data. The final classification consists of 10 Groups and 53 Types.

7

EVALUATING GEODEMOGRAPHIC PERFORMANCE FOR PROFILING OF ACCESS

7.1 Evaluation of Discrete Classification

It was argued in Chapter 6 that when creating geodemographic classifications the division of areas into clusters of homogeneity can have numerous outcomes through altering the numbers of clusters, the number of hierarchies in the typology, the geographic scale used to create the clustering units, and finally the multiple possible arrangements of spatial clustering units within an "optimised" cluster model. The lattermost of these issues relates to how the *k*-means algorithm can have multiple locally optimised models depending on the location of the initial seeds used as input into the cluster analysis. Chapters 3 demonstrated how there are numerous classification types used in applications both across the public and private sectors. However as Chapter 6 identifies, none these is designed for public sector applications, nor more specifically the domain of Higher Education. The use of geodemographic classifications thus far within this thesis has remained uncritical; however this chapter aims to evaluate the performance of a number of discrete models for Higher Education data segmentation using a series of evaluation methods.

The evaluation of discrete classifications can be considered both quantitatively and qualitatively. A qualitative evaluation focuses on those attributes which are considered to make one classification more fit for purpose over another. Indeed, Leventhal (1995:6) argues that "while numerical measures provide helpful summaries, they cannot evaluate the usefulness of a discriminator as opposed to the statistical significance". Qualitative evaluations could be conducted for a particular application (e.g. creating targeting strategies to increase a course's recruitment), or as an overall assessment for use of a classification within a particular sector (e.g. HE). These types of evaluation are subjective, and as such should not be used in isolation to determine the suitability of a classification. They nevertheless do provide a good framework to learn about the numerous discrete classifications which may be suitable, and also provide a method of choosing which classifications could be used in more rigorous empirical testing.

The data used for the quantitative evaluation are derived from 2004 UCAS acceptances for England to enable the broadest range of classification to be compared. Using a Total Weighted Deviation (TWD) method, the predictive performance of classifications may be examined for courses of Higher Education and also for aggregate institutional profiles. These represent two core analyses that an institution might complete in order to market or extend participation to people living within specific neighbourhoods. A second evaluation method compares the performance of the classification to identify and therefore target neighbourhoods which are unlikely to supply traditional aged (18-19) participants to study degrees in Higher Education. The classifications compared in this evaluation are those which were made available to the researcher which included the National Statistics Output Area Classification, Experian's Mosaic, CACI's ACORN, the bespoke classification created in Chapter 6 and the Index of Multiple Deprivation (IMD). The IMD is measured on a continuous scale, however for the purpose of this analysis the classification was divided into 10 deciles. For the TWD analysis only the National Statistics Socio-Economic Classification (NS-SEC) and the HFCE Participation Groups (POLAR) are also evaluated. They are excluded from the aggregate participation performance analysis as 18-19 year old base scores could not be derived for a comparable time frame.

7.2 Qualitative Analysis: What Makes a "Good" Geodemographic Classification?

The first stage in a qualitative evaluation is to identify a set of criteria through which the classifications will be assessed. These might include:

- Frequency of clusters
  - More clusters may be considered advantageous as this would provide a finer level to neighbourhood profiles.
- Hierarchy of clusters
  - An increased number of hierarchies demonstrate that the classification has flexibility when analysing target groups of varying population sizes. Partitioning a small population by a typology consisting of many categories would create unreliable / unstable index scores.
- Geographic scale of classification
  - A classification which identifies neighbourhoods (clustering units) at a fine scale such as the individual, household or unit postcode is, all other things being equal, of greater usefulness than one available only for coarse aggregations.
- Suitable labelling of clusters
  - The clusters have memorable names which communicate their key attributes to a target audience.
- Suitable "Pen Portraits" to describe the clusters
  - Full contextual descriptions on the attributes which describe people living within the neighbourhoods defined by the clusters.
- Appropriate multimedia and imagery

is better interpreted as a method of deriving the strengths, weaknesses, opportunities and threats for a particular segmentation solution prior to more detailed quantitative assessment of the discriminatory power of each classification.

7.3 Methods of Quantitative Evaluation

Quantitative evaluation will be complete using a total weighted deviation (TWD) method to examine the predictive power of classifications across a range of variables; and Lorenz Curves (Gains Curve) and Gini Coefficients to examine a single application of predicting the participation rates of young participants to study degrees. In order to create a measure of young participation as a function of the underlying base population, the estimated frequency of 18-19 year olds from the census was required at postcode level. Data from the 2001 Census are distributed at its finest level within Output Areas (OA), which although appropriate for OA level classifications such as OAC, others such as Mosaic and ACORN classify neighbourhoods at the level of unit postcode. An additional output from the 2001 Census available from the Office of National Statistics (ONS) is a postcode file which disaggregates the total, male and female populations by unit postcodes. Using the distribution of the unit postcode level total population frequency (*R*) within an Output Area (*OA*) the 18-19 year old data (*T)* could be apportioned, thus estimating frequency of 18-19 year olds at postcode level (*Y*) based on an assumption that this demographic is distributed across each OA uniformly (see Equation (7.1)). Because these data would later be aggregated into the classification categories, the data were not rounded, and as such a unit postcode could apparently contain non-integer values with respect to 18-19 year old persons.

$$y_i = T_{OA} \times \frac{R_i}{\sum_{i=1}^{n} R_i}$$

(7.1)

Apportioning population in this way is computationally intensive because of the size of the files being manipulated. The apportioning algorithm was run iteratively for every OA using the statistical software SAS, taking a number of days to run through to completion. This type of analysis requires a significant level of statistical and programming ability and as such is not an analysis that many end users of geodemographic classifications would be either able, willing or have time to complete. For specific applications dealing with a sub population group this makes OA level classifications far easier to use than those based on postcodes. To circumnavigate these issues commercial vendors will often supply a software application where these analyses are predefined.

7.4 Social Similarity, Clustering Scales and Indices of Dissimilarity

Chapter 6 presented an argument that attribute space can be partitioned in multiple arrangements and therefore that geodemographic classifications do not present a definitive or uncontested representation of reality (Gordon, 1980). However, despite these multiple possible cluster solutions, it appears that geodemographic classifications do share common characteristics of social similarities within the groups of classification typologies, that is clusters have shared characteristics within their common attribute space. Table 7.2 shows index scores and percentages which demonstrate the degree of similarity between Mosaic Groups and OAC Super Groups. While a number of Mosaic Groups have significant overrepresentation within particular OAC Supergroups, others show less extreme overrepresentation with a more even distribution. A commonality between those which are more evenly distributed is that they are predominantly coded by Mosaic as people living in more affluent areas. If reality is assumed to be accurately represented by the Mosaic Classification, it could be interpreted that these more affluent neighbourhoods are less well defined in the OAC classification, i.e. by the census. However, what is encouraging for those using OAC in applications targeting less deprived areas (as extracted from the Mosaic descriptive material), is that this classification shows a great deal of similarity to Mosaic across these neighbourhoods.

scale, such as OA, small changes in the population will not alter how an area should be most appropriately classified, as the average characteristics of those resident still within the area will predominantly remain aligned to the classification representing these aggregate characteristics. Furthermore, those micro-level data used by commercial classification builders often originate from sample surveys which are not representative of the total population, and as such the implied precision of a classification at unit postcode level accuracy may be superfluous. Indeed, the use of these data may explain why in a minority of OAs in this analysis were found to be very heterogeneous, with one OA in Northern Ireland containing 7 different Mosaic Groups, perhaps caused by erroneous imputation of sample data or the unavailability of household level databases. Finally, when Mosaic Groups were compared with OAC Supergroups there was a high degree of social similarity between the classifications indicating comparability between how the two classifications categorise neighbourhoods. These effects seem to occur despite the modifiable areal unit problem (MAUP) (Openshaw, 1984) of scaling (Wrigley *et al*, 1996) between the postcode and Output Area. However, these similarities were less pronounced in the more affluent Mosaic Groups, and may indicate a possible weakness in OAC as these neighbourhoods would be more difficult to identify.

7.4.1 Indices of Dissimilarity

A further method by which the heterogeneity within OAs can be gauged is through the use of diversity indices. These are established techniques in ecology to measure the heterogeneity of species within biological areas such as transects. These techniques can be adapted to provide a further method of measuring the homogeneity of Mosaic Groups within OAs. The Simpson Diversity Index (Simpson, 1949) measures the balance of neighbourhood groups within an OA and the scores range between 1 and 0. A more heterogeneous OA would contain a number of different Mosaic Groups and have a score closer to 0, whereas a homogeneous OA would have fewer Groups and a score closer to 1. The Simpson Diversity Index (D) in OA e is calculated by summing the square root of m Mosaic Groups ranging from i=1 to =11 which correspond to each of the 11 Mosaic Groups (See Equation (7.2)).

$$D_e = 1 - \sum_{i=1}^{11} \left(m_{ei}^{\ 2} \right)$$

(7.2)

Once scores have been created for each OA they may be mapped to examine the degree of OA level homogeneity (See Figure 7.7 and Figure 7.8). It can be seen that in urban areas there is a tendency for OAs to have a higher diversity (low score) with rural areas such as North Devon, Mid to North Wales, North Lancashire and the majority of Scotland all having high scores, i.e. low diversity. One exception to this broad pattern is that Northern Ireland appears quite diverse, even in areas which would be considered predominantly rural. Using the OA scores it is possible to aggregate the diversity scores into the OAC typology in order to assess whether particular clusters demonstrate heterogeneity within their assigned OA. Because OAC clusters vary in size, these aggregated scores were transformed into index scores using a base of the total OAC distributions. Thus, Figure 7.9 represents those neighbourhoods by the OAC classification which typically show a greater heterogeneity at unit postcode level within the OA. A high index score indicates an area which is heterogeneous, and a low index score more homogeneous. There are differences both between and within the Super Groups and the following section presents a number of interesting findings. The Super Group "Constrained by Circumstances" contains three Groups. The first ("Senior Communities") and last ("Public Housing") of these Groups are both more homogeneous while the middle Group ("Older Workers") are more scattered alongside other Groups within OAs. From within the Super Group "Blue Collar Workers" the last Group "Older Blue Collar" are predominantly more heterogeneous indicating a demographic dimension which relates to this diversity at postcode level within these areas. Within the Super Group "Countryside", the Group "Agricultural" is shown to be more homogeneous. The other two groups within this Super Group are areas which are more urban such as small villages or towns. It is hinted in Harris *et al* (2005) that Experian may re-cluster non urban areas in a separate analysis in order to create clusters which represent a finer level of detail. This could possibly explain why the Groups "Village Life" and "Accessible Countryside" show greater heterogeneity. The

Equation (7.4) calculates a q_{ij} predicted rate for an i category by taking the n total postcodes recorded for each j variable (E.g. Course Grouping, ethnicity etc) and multiplying by p_i across p_1, p_2, ..., p_n.

$$q_{ij} = n_j \times p_i$$

(7.4)

All q_{ij} predicted scores can then be substituted from the t_{ij} actual recorded frequencies to give the difference between the predicted and the actual counts as r_{ij}.

Once a set of r_{ij} scores is created for j_n variables across a discrete classification a total deviation (TD) statistic can be derived. A standardised statistic can be calculated across the differences (r_{ij}) to measure the dispersal of scores from the mean. The mean of this standardised statistic is 0 for all variables because the discrete classifications categorise 100% of the population, and where a discrete classification does not have a category for a unit postcode these are classified as "Unknown". Thus, the calculation to derive TD where p is the frequency of categories for the discrete classification is shown in Equation (7.5). The $(p\text{-}1)$ correction in the denominator is not used as the calculation refers to the total population and not a sample (Hardy and Bryman, 2004).

$$TD = \sqrt{\sum_{i=1}^{n} \frac{{r_{ij}}^2}{p}}$$

(7.5)

Total deviation statistics calculated for large populations should be considered more reliable as they are less vulnerable to 'freak' patterns caused by outlier scores. Population weighting circumnavigates this potential problem, so the TD scores are multiplied by a proportion created by dividing the sum of the target variable frequency by the total base frequency (See Equation (7.6)).

$$TWD = TD \times \frac{\sum_{i=1}^{n} t_{ij}}{\sum_{i=1}^{n} b_i}$$

(7.6)

TWD is thus a measure of classification performance on a linear scale which accounts for a variable's population size in its derivation. The technique is suitable for comparison of variable sets across a number of classifications, and thus provides a method of evaluating classification performance where a larger TWD is attributable with better discrimination between areas. Thus, for the comparison in this chapter these TWD measures of predicted versus observed postcode frequency were calculated from UCAS 2004 acceptance data for courses and institutional profiles. The UCAS data were aggregated by the following classifications:

- Mosaic Groups (MGroup)
- Mosaic Types (MType)
- ACORN Category (ACat)
- ACORN Group (AGroup)
- ACORN Type (AType)
- POLAR (POLAR)
- Index of Multiple Deprivation (IMD)
- National Statistics Socio Economic Classification (NS-SEC)
- OAC Group (GROUP)
- OAC Subgroup (SUB)
- OAC Supergroup (SUPER)
- Educational OAC Groups (EDUOACG)
- Educational OAC Types (EDUOACT)

Educational OAC Groups appear marginally above the line; however the explanatory powers (TWD) of these classifications vary widely within this range. ACORN Types is the highest performing classification, although Mosaic Types, OAC Subgroups and Educational OAC Types also perform very well. OAC Subgroups and Educational OAC Types appear to perform quite similarly. It would have been desirable if the performance of Educational OAC Types had increased by a great proportion over OAC Subtypes, however as the build methodology was directly linked to OAC, this may have had a bearing on the performance gain. The increased performance experienced by the unit postcode level classifications of ACORN Types and Mosaic Types over those which classify at OA could be related to these classifications using a finer geographical scale. Thus, there may be argument to suggest that an Educational classification should be built at the scale of unit postcode. However, at Group level, this advantage appears eroded with Educational OAC Groups only marginally underperforming Mosaic Groups.

7.6 A Lorenz and Gini Coeffient Evaluation of Youth Participation in Higher Education

A further evaluation was conducted using a Lorenz curve (Lorenz, 1905) which illustrates the gain in discrimination one would expect by using a particular classification over the null hypothesis (straight line) that all areas supply participants to Higher Education in equal proportions. The larger the area under the curve, the greater the total discrimination, where a theoretical maximum score would be equal to one. These area scores were calculated using a Gini coefficient (Gini, 1912) as shown in Equation (7.7) where *x* are the 18-19 year olds in cluster *k* and *y* are the 18-19 year old participants in cluster *k*.

$$G = \left| 1 - \sum_{k=1}^{n} (y_{k-1} + y_k)(x_{k-1} - x_k) \right|$$

(7.7)

The Lorenz curves for a series of geodemographic classifications are presented in Figure 7.11 and the rank order of their Gini coefficients shown in Table 7.6 and Figure 7.12.

rate to be predicted by a classification (as measured in the Gini Coefficent analysis), this is a less useful application than, for example, predicting a particular course participation propensity (as measured in the TWD analysis). With the advent of national educational databases such as those accessed for use in this thesis, there is little need to predict aggregate participation when actual participation rate can be derived and mapped (e.g. POLAR). As was previously shown in Chapter 4, participation is correlated with "wealth" and classifications that stratify neighbourhood Types by this attribute will demonstrate a stronger prediction of participation, as has been shown by the performance of the Mosaic classification in the Gini Coefficient analysis. However, the TWD analysis demonstrated that when disaggregating participation into course and institution categories, Mosaic performed less well than might be expected, being out performed by ACORN Types. One possible interpretation of this finding is that the Mosaic classification may stratify too highly by measures of affluence, and not enough by other attributes which may influence or be important for measuring disaggregated applicant behaviours to participate in particular courses or institution.

7.7 Conclusion

This chapter has shown how geodemographic classifications can be evaluated both quantitatively for their relative discriminatory power and qualitatively to assess classification features against user requirements. Cross tabulations between classification typology have shown neighbourhoods possessing similar characteristics from with the typology. Through the comparison of OAC with Mosaic, it was also possible to see which Mosaic Groups are formed from data related to income data, which are absent from OAC. A detailed study of the relative information loss one can expect from switching between a classification that categories neighbourhood at OA rather than unit postcode was examined. On the assumption that Mosaic represents an accurate depiction of reality it was found that OA diversity was generally higher in urban rather than rural areas. Furthermore, the patterns of diversity were more prevalent for OA categories by particular OAC Subgroups which indicates where the additional data used to construct Mosaic provide extra discrimination, or, that through weighting schema particular characteristics have been created in the Mosaic classification. A series of classification were compared using a total weighted deviation and Gini Coefficient / Lorenz curve. These measure the ability for classification to account for variations in participation rates, and disaggregated participation behaviours at course and institution level. Possible issues with the way in which the Mosaic typology stratifies neighbourhoods were highlighted in this analysis.

PART 4: ARE THINGS GETTING BETTER?

8

TOWARDS A MORE MERITOCRATIC MARKET?

8.1 Introducing Temporal Access Change

Chapter 2 describes the history and development of mass participation in Higher Education from medieval times through present day. Although an expanding sector provides advantages to more people, this thesis is concerned with the extent that these life chances are equally distributed across society and space. Chapter 4 established the existence of those inequalities current policy initiatives are designed to address and this chapter extends this analysis to further examine temporal trends in access inequality and participation, both empirically and through critique of how these activities are funded. Early geodemographic classifications were updated relatively infrequently using predominantly decennial census data. However, commercial classifications in 2007 are privy to greater sources of data which are refreshed more frequently. Similarly, in the public sector data are collected annually, however unlike many of the commercial datasets which often relate to small samples of the total population, the public sector data are often far more comprehensive in target population coverage. The rate of change in the Higher Education sector in particular is very rapid, and as such it makes sense to consider the longitudinal profile of the recent past.

(1996: 385) describe performance indicators as "a summary statistical measurement on an institution or system which is intended to be related to the 'quality' of its functioning", which in the case of the HESA performance indicator "quality" is held as a relative and standardised measure.

Perhaps a more appropriate method of evaluating changes over time could be a refined version of league tables (see Section 2.5.1), like those composed each year by many of the national press (E.g. The Times Good University Guide[50], Guardian University Guide[51]).League tables take a different view on "quality", and one in which institutions are compared and ranked by their merits or failures across a broad range of assessment criteria and dislocated from any contextual factors. It is usual for national averages to be used for these comparisons as the overall aim is typically to create a ranking above or below this point. In the context of extending access these types of measures can be used to demonstrate which institution or course type are:

- Increasing the frequency of admitted students from low participation backgrounds.
- Increasing the frequency of admitted students from low participation backgrounds relative to institutional growth.
- Increasing the frequency of admitted students from low participation backgrounds relative to institutional growth and changes in the background frequency of these groups across society.

Therefore, using various UCAS data from 2001 – 2006 this chapter aims to explore both absolute trends in participation and widening access and also disaggregated trends within institutions and course groupings. The temporal period considered within each analysis vary depending on the availability of data or variables for a

[50] http://www.thegooduniversityguide.org.uk/
[51] http://education.guardian.co.uk/universityguide/0,,488282,00.html

particular period in time. Adjusted benchmarks will therefore not be used in these comparisons as the aim is to assess the absolute rather than standardised (relative) differences and to demonstrate the true stratification of access over time, rather than measure standardise performance at improving access inequality.

As discussed in Section 2.3 there are multiple definitions of Higher Education and one way of demonstrating that participation has improved is to increase the definition of which courses of study (and therefore people) are included within "Higher Education". The 50% national participation target discussed above is based on students aged between 18 and 30, and include any course which is above an A-Level (i.e. Level 4+ in the national qualifications framework) and leads to a qualification which is awarded by a Higher Education institution or national awarding body (Prospects, 2002). This is a broad definition of Higher Education and is not adopted in the benchmarks produced by HESA who manage statistics for a more compact range of institutions. The analysis presented in this chapter concern those who have accepted full time degree courses at UCAS institutions.

8.2 Accounting for Higher Education Growth

As shown in Chapter 2, the Higher Education sector has continued to expand and increasing participation towards a 50% participation target has continued to occur. However, the extent that this growth has occurred evenly across all subjects can be investigated using UCAS data from 2002 – 2006 (see Table 8.1 and Figure 8.5). This table shows that there are course groups which have grown by as much as 39.8% (D Veterinary Science, Agriculture & Related) where as others have declined by 24.7% (Non-European Languages & Related). There are too many subjects to analyse individually so a number of the subject Groups are examined in more detail at Line level in Sections 8.2.1 - 8.2.3.

8.3 Widening Participation Profiles Over Time

In Section 2.1 it was shown that changes in the proportion of students attending Higher Education from different social class groups had remained relatively static since 1968. However, as was discussed in Section 3.4 these measures are increasingly unreliable because of the fluidity of the current job market and the assignment of the groups based on parental occupation. It was then argued that geodemographic analysis provides a more robust measure of socio-spatial stratification. Before temporal rates of neighbourhood access are considered, a point raised in Section 2.1 will be addressed. This was an increasing trend for accepted applicants to be recorded by UCAS with a current occupation classification of "unknown".

8.3.1 Social Class and the "Unknowns"

One measure of increasing access is the frequency of students who accept places on degrees by their National Statistics Socio Economic Classification (NS-SEC). The frequency scores were obtained from the UCAS website for 2002 – 2006 and are shown in Figure 8.2 as prior to 2002 a different socio-economic classification was used. In terms of absolute numbers there has been little change in the frequency of students between groups; however there has been a growing trend for the increase in students classified as "unknown".

In order to look at the relative access rates between years a measure of the change in base employment characteristics over the period 2002 to 2006 was required as this has an effect on the supply of different people into Higher Education from backgrounds classified into NS-SEC. The largest sample data available were found in extracts from the National Statistics Labour Force Survey (LFS)[58]. The 2002 - 2006 base data were combined with degree acceptances counts over the same time period and index scores were calculated (see Figure 8.3).

[58] http://www.statistics.gov.uk/STATBASE/Source.asp?vlnk=358

Index changes in the bottom 10 institutions are all reasonably low however, there are some excellent results in the top 10 institutions that have extended access greatly. An index score change of 26.6 at University of Sunderland and 24.1 at the University of Greenwich were achieved, where both institutions experienced growing student numbers.

It is difficult to evaluate the success or failure of widening participation initiatives at an institutional level as a function of the funding received from HEFCE. These monies are vested in institution to fund a variety of activities; some may result in students being recruited directly into the institution as a function of widening participation activities, whereas some of these activities may generate spill over effects whereby the students targeted by one institution's activities may choose to attend a different institution, but still to the benefit of widening participation in the Higher Education sector as a whole. However, within these caveats the remaining part of this section presents exploratory analysis which compares the widening participation funding received for English institutions when compared to the recruitment of students from a widening participation background. This analysis is limited to English institutions funded by HEFCE because of restricted availability of temporal funding data[59] and where a match was possible with UCAS data for the same time period. The definition of widening participation students remained as those from neighbourhoods identified by the ACORN Category "Hard Pressed", and only the funding allocated to full time study was considered.

The calculation of how these funds are allocated has been outlined by HEFCE (2007c). An institution receives funds based on the 3 step process outlined below:

- Step 1 – Each full time undergraduate attending the institution in the previous year is mapped to a 1991 Census Ward where they are assigned into one five POLAR quintiles relating to the aggregate ward participation

[59] The temporal funding data was provided by HEFCE through personal contact with their analytical services department.

effectively demonstrated by the creation of the "Center for Geographical Analysis" in 2006 at Harvard University (Gehrman, 2006). This centre focuses on GIS, which Harvard University is calling the "new Geography". This direct comparison to the US is perhaps a little unfair, as in the UK *new* Geography includes a plethora of study areas such as the geographies of car interiors (Ashton, 2005), otter hunting (Allen, 2005) and pet therapy (Kearns, 2005). Although these are a bias selection of topics covered in the talks from the RGS 2005 conference, they do highlight how fragmented the discipline has become in the UK. "Geography is the study of the Earth's landscapes, peoples, places and environments" (RGS,2007), and it is argued here that this literal meaning should perhaps be replaced by something more appropriate that addresses key issues and problems at a scale where solutions or interventions can be effectively implemented.

The increased interest in Forensic Sciences generated by television programmes should provide evidence to the Geography advisory committee on how the student market responds to targeting initiatives, and although curriculum based initiatives may raise overall subject teaching quality, there is no guarantee that this will raise popularity. Enlisting the help of Michael Palin for the launch event will help, however the Action Plan for Geography money may have been better spent on creating learning resources around television programmes with a geographical focus, or even enlisting promotional or marketing experts to reignite interest in the discipline amongst school children through initiatives which are demographically appropriate. The BBC partnership with the Open University is a very good example of how Geography education can be linked to television shows such as Coast through online[62] and offline course material and resources, and it is these activities which should be encouraged.

Perhaps what should be of more concern to Higher Education Geography departments than a declining popularity are the apparent increasing concentrations of students from more affluent areas in the overall student demographic. Between

[62] http://www.open2.net/coast/

students were most and least likely to participate in Higher Education. Rates of change of this period were also shown to be heterogeneous between courses and institutions. Current HEFCE policy relating to how widening participation funds are allocated was evaluated in relation neighbourhood participation rates, both in absolute terms and how these rates have changed over time. It was concluded that current policy focus on funding per head of participants from a widening participation background (defined by POLAR) may ignore the reality that some institutions recruit these students very easily, whereas others find it more difficult. It also identified that there is no mechanism within the current funding model to reward institutions who have improved their access rates over time. The final part of this chapter examined extra funding initiatives which are additional to HEFCE widening participation funding to address particular subject concerns such as the declining popularity of Geography, both in schools and Higher Education. The Geography Action Plan was highlighted as raising a number of concerns, particularly over the choice of schools that were targeted. These issues perhaps arise because of a lack of centralised information service for those involved in educational decisions making.

9

A GALLERY OF APPLICATIONS FOR HIGHER EDUCATION STAKEHOLDERS

9.1 Higher Education Stakeholders

Throughout this thesis arguments have been built which suggest how a range of rich data pertaining to participation and performance in education can be integrated to provide both cross sectional and longitudinal information through relevant and bespoke indicators. For those stakeholders in Higher Education, that is students, schools and universities, these analyses help exemplify how more informed decisions can be made. Schools are seen as an important link in this chain as they contain those student "customers" of Higher Education, and thus schools need to be equipped to give suitable advice depending on students' needs and background. This chapter therefore presents a series of applications which utilise some of those techniques employed throughout this thesis and integrates them within broader GIscience techniques to investigate school "market areas".

9.2 A Regional Case Study – Stakeholders in Manchester

For this regional case study in Manchester, Higher Education stakeholder data is analysed to examine how patterns of socio-spatial differentiation in access to Higher Education are created in part as a function of local processes occurring in schools and colleges. The local authority of Manchester provides a useful analytical area as it contains a mix of school types with variable attainment.

9.2.1 Prior Attainment

Earlier in this thesis Section 4.4.1 showed the general trends in pre and post 16 neighbourhood level attainment without reference to local school effects. Students progressing into study at Post 16, possess qualifications from a variety of schools, each with a different mix of students and attainment. The following analysis examines the interaction between neighbourhood attainment and school attainment in Pre 16 schooling to demonstrate how prior attainment is spatially heterogeneous, thus creating an early filter on those students who will progress into Further and Higher Education. The average GCSE point scores were calculated for the neighbourhood groups within schools and compared against the average school attainment to analyse which neighbourhood groups within the schools were contributing most to the overall attainment. A further comparison was made by comparing the average attainment for neighbourhood Groups within a school against the same Groups in the whole of England to benchmark school performance in terms of raising the attainment for students from specific neighbourhood Groups. In this analysis the attainment for neighbourhood Groups within a school were only analysed where there were more than 5 students within the Group to prevent average scores being taken from very small numbers of observations.

These comparisons are illustrated for two of the schools within the local authority of Manchester. The geodemographic profile of Parrs Wood High School for those completing GCSE at the end of KS4 is shown in Figure 9.1. Figure 9.2 shows how the attainment from these different neighbourhood Groups compared to the school average. Those students attaining the highest GCSE scores are those from the more affluent areas. In all of the neighbourhood level attainment graphs that are presented for schools in the following section, where the number of pupils within a group was 5 or fewer then the results were suppressed. The numbers which label the columns are the frequencies of students within each neighbourhood group.

although a school may show a high profile for low students from participation neighbourhoods, these students may be relatively poor in attainment.

9.2.2 Spatial Distribution of Advantage – Overlapping Markets?

The previous section has shown how attainment at KS4 and KS5 is stratified between schools and spatially by neighbourhood type. However, stakeholders may also be interested to know why these neighbourhood patterns exist, and in order to explore these issues a method of visualising the geographic distribution was required. Thus, educational geodemographics can exploit the toolkit of GIS/GIScience which has long been used in retail studies of "market areas" (Birkin *et al*, 2002). Although an average score of the distance that students travel to study A-Levels is a useful measure of the extent of a school/college attendance area, this neither demonstrates where geographically the students are supplied from, nor measures the competition between school areas of acceptance. Thus, in order to examine the spatial patterns of school recruitment, a series of maps have been created using a kernel density estimation technique (De Smith *et al,* 2007). This technique measures the density of students within a kernel bandwidth (e.g. 1000m) and outputs a raster grid at a prespecified resolution (e.g. 250m) containing the average density of students. Additionally, using the implementation of this technique in Hawths Tools[65], an extension for ArcGIS[66], it is possible to extract a vector boundary delineating different volume thresholds (the per-cent volume contour, PVC). Such features can be used to identify the area which likely contains a given percentage of the students attending a given school or college. Gibin *et al* (2007) illustrate this as shown in Figure 9.10, using a density surface for two points. The horizontal plane through the two kernels centred upon the two points represents the 50th percentile, and is the point at which the PVC area and shape are calculated. The PVC is represented in the left hand diagram using a white line.

[65] Hawths Tools can be downloaded for free from: http://www.spatialecology.com/htools/
[66] Arc GIS is a software package available from ESRI : http://www.esri.com/software/arcgis/

responses to a Nelson Non Verbal Reasoning Test[69] (NVRT) which is used as part of this school's selection criteria. The criteria used to allocate places at this school are as follows[70]:

1. Students with Statements of Special Educational Needs where the College has consented to be named in the Statement.
2. Students in public care (looked after children).
3. 10% of students are admitted on the basis of aptitude in Music, using a specified assessment process.
4. Students for whom it is essential to be admitted to the College because of special circumstances to do with significant medical or social needs evidenced by written medical advice.
5. Students who, on the date of admission, have a sibling (i.e. a natural brother or sister, or a half brother or sister, or a legally adopted brother or sister or half-brother or sister; who will be living with them at the same address at the date of their entry to the College) on the roll of the Haberdashers' Aske's Hatcham College.
6. Of the remaining places :-
    A. 50% are offered to students living within three miles (4.8Kms) and south of the Thames, on the basis of proximity; i.e. students who live the nearest radial distance to the College on the close of the admission application date.
    B. The remaining 50% are offered to students living within three miles (4.8Kms) and south of the Thames, on the basis of an independently operated random allocation.

[69] The Nelson Non Verbal Reasoning Tests are available from: http://www.nfer-nelson.co.uk/
[70] Source: schools admissions criteria: http://www.haaf.org.uk/hpropectus-admissions/hprospectus/hadmissionscriteria.htm

The national adoption of randomised selection criteria requires further analysis and modelling to ensure that attainment and quality within the sector do not fall. Further investigation is needed into which are the most appropriate methods of measuring students latent ability, how many bands should this ability be divided into and the frequency of students that should be placed within them. Finally, from a policy discourse perspective, "piecemeal social engineering" (Popper, 1971) which restricts a school's ability to create their own selection criteria should be examined against other free market models of school place allocation such as voucher schemes.

9.5 Conclusion

This chapter develops a series of case studies relevant to a range of stakeholders in Higher Education. These directly relate to a series of analysis that these groups of people may wish to perform including

- Exploring the effects of different school selection policy
- Investigating operational issues such as selecting the most appropriate schools to target with widening access initiatives
- Performing geographical analysis of school market areas
- Creating school performance benchmarks to target resources to raise attainment in low participation neighbourhoods

Furthermore, GIS/ GIScience functionality has been shown to be relevant to extending educational geodemographics to visualise geographic market areas of schools.

10

CONCLUDING COMMENTS

The purpose of this thesis has been to investigate how best to represent the multiple dimensions of social, spatial and temporal processes which shape access to Higher Education in the UK. Full and contextualised understanding of these concepts is seen as acutely important to stakeholders (pupils, universities, schools) in Higher Education. Recent changes in the policy and funding of Higher Education have presented institutions with the following challenges:

- To demonstrate how sector and institutional data might enable universities and colleges to position themselves within a changing policy framework
- To understand how the information that institutions themselves collect may be best utilised to help institutions design their marketing and student support strategies
- To identify the geographical areas from which institutions should be able to attract students
- To analyse the nature of the disciplines offered within institutions and the types of students they attract

The agendas of widening participation, extending access and institutional marketing share common challenges to devise better ways of reaching potential students who are appropriately qualified and motivated to pursue and successfully complete the full range of institutional course offerings. One key motivation for this thesis has been the observation that few Higher Education institutions have communication strategies that are tailored towards reaching the full range of potential students who would benefit from their current subject and course offerings. University and college marketing initiatives are often unsystematic, even serendipitous, in the ways in which they identify schools and colleges for outreach and widening participation initiatives, and sometimes uncoordinated in the ways in which they present the full institutional profile of subjects of study in these activities. Thus, a core objective of this thesis has been to set out some relevant aspects of the changing Higher Education policy-setting arena and to present a systematic framework for widening participation and extending access in an era of variable fees. In particular the thesis aimed to illustrate how Higher Education data and publicly available sources might enable institutions to move from piecemeal analysis of their intakes to institution wide strategic and geographically linked market area analysis for existing and envisaged subject and course offerings.

The first part of this thesis demonstrated how the current Higher Education system has evolved over a period of growth during the last 50 years. Historically it was suggested that this system has served the privileged minority, although contemporary Higher Education was seen as increasingly being transformed into a system which provides for the mass market. Through the introduction of student fees Higher Education is rapidly developing the characteristics of a traditional commercial market, thus increasing demand for information by both consumer and producer stakeholders. Despite rapid growth in the absolute numbers of students attending Higher Education, it was suggested that there remain access inequalities between students from different socio-economic backgrounds. Monitoring of the Higher Education sector was shown to involve numerous organising bodies, each with their own remit, data and collection mechanisms: the result was described as a current educational data economy characterised by overlaps, lack of coordination and missing data. It was outlined that there was little in the way of the kinds of centralised and customer orientated services which are essential and relevant for making decisions in Higher Education. Thus, an important contribution of this thesis is that it is one of the first pieces of work to systematically attempt to integrate and bring these data together for the benefit of Higher Education stakeholders. The final contribution in the first part of this thesis was to highlight that there remains a lack of clear thinking about how concepts of social capital map into the quantitative variables available which permit generalisation, and that there is a need for composite multivariate indicators that are robust, transparent and safe to use.

The second part of this thesis demonstrated how stakeholders in Higher Education are currently making uninformed choices in those areas which are of acute policy concern and there remains a need for much improved decision support tools. It highlighted a need to measure and monitor change in this fast developing field even more so than in the traditional applications domains of geodemographics. This part of the thesis began with an exploratory data analysis which demonstrated that access to Higher Education is both spatially and socially complex. The distance that applicants travelled to accept courses of Higher Education was related to the specialism of a course, the type of institution to which they were applying, and the tariff score required of the applicants. Higher Education qualifying attainment scores were shown to be differentiated by neighbourhood Type, and higher attainment was predominantly observed in students from more affluent neighbourhoods. Those national indicators and benchmarks used to assess performance and allocate funding to institutions for widening participation were critically examined. These were shown to be of variable quality in terms of identifying the neighbourhoods least likely to participate in Higher Education. The complexity of the education sector was illuminated in these analyses and highlighted a critical need for better information and decision support tools. A pilot Educational Market Profiler was introduced in response to some of these shortcomings. This was designed in collaboration with UCAS to extend their services to provide centralised and freely available profiling information to their partner Higher Education institutions. This tool was created from a user requirements analysis and also linked to the theoretical concepts and empirical observations presented in earlier parts of the thesis. A case study was developed with University College London to demonstrate how the analysis produced by the tool could be used in "real world" problem solving and decision making. A unique technical contribution of the Educational Market Profiler was to demonstrate a unique methodology for the integration of data from the Higher Education sector and schools, enabling profiles to be built for both those who attend, and those who do not attend Higher Education.

The third part of the thesis investigated how the classification methods used in this thesis could be built upon to provide more relevant indicators for Higher Education applications, and suggested a series of methods through which discrete classification can be evaluated against these aims. The addition of Higher Education data into a geodemographic classification with open methodology was seen as a positive step beyond the use of generic and re-labelled commercial classifications, and was seen as an improvement of use of classifications for purposes beyond those for which they were not originally designed. This provided a challenge to the implied assumption that individuals use public services such as Higher Education in ways directly analogous to the ways in which consumers use private goods. These analyses have demonstrated the first bespoke geodemographic classification created specifically for a Higher Education application using public sector data. A further technical and computational contribution included exploratory analysis of how the output from a *k*-means clustering algorithm might be made more robust through re-running models multiple times to extract an optimised solution. Evaluation of geodemographic classification demonstrated some of the unique and shared properties between the commercial classification Mosaic and OAC. Through these comparisons it was possible to see which Mosaic Groups are formed from data related to income, and the relative information loss that one might expect from the switch between a classification that categorises neighbourhoods at OA rather than unit postcode scale. Using a series of evaluation methods the performance of classifications were compared, in order to effectively partition both course and institution access rates and to predict levels of aggregate participation. Although the commercial classification demonstrated marginally better performance, possible issues with the way in which the Mosaic typology stratifies neighbourhoods were highlighted in this analysis.

The final part of the thesis examined whether the inequalities highlighted in this thesis are getting better or worse and how these can be addressed through engaging with stakeholders in Higher Education. These analyses extend from previous analysis to examine longitudinal trends related to participation rates and educational histories of potential applicants. These discussions relate to a need to theorise the links between supply and demand in order to suggest routes towards more efficient and effective distribution of the life chances of students, in order better to accommodate their likely needs, preferences and the uniqueness of places within generalised representations of the system. It was shown how the Higher Education sector has variable growth rates between courses within institutions, and that these changes are not evenly distributed across societal groups. The thesis further illustrated how socio-economic status, which is relied upon for much sociological and educational research in the area of access to Higher Education, is slowly being eroded as a useful indicator through increasing rates of incomplete data. Using neighbourhood classification access to Higher Education from low participation neighbourhoods was shown to have improved marginally over the period 2001 – 2004, although strong inequality in participation rates between those neighbourhoods most and least likely to attend Higher Education still remain. The final investigation presented in the thesis was a series of analysis case studies relevant to a range of stakeholders in Higher Education.

Within this concluding chapter it is also appropriate to summarise some of the limitations of the methods employed in this thesis. From the preceding review and analysis, it is possible to identify three main issues of uncertainty associated with the use of Geodemographic analysis as an organising framework. These concern the statistical stability of Geodemographic profiles, the effectiveness of Geodemographic indicators in capturing the multivariate influences on Higher Education participation, and the uncertainties that geographic variation may introduce in any representations made.

Geodemographic analysis has been shown in this thesis to provide useful descriptive summary measures for use in Higher Education applications. However, the statistical significance of differences between index scores was not ascertained. Such analysis would make it possible to quantify the probability that observed differences might be attributable to sampling variation. When using other established statistical methods (such as regression analysis, for example), associated goodness of fit statistics may be used to measure how well a model replicates the results of a dataset. Such statistical techniques could usefully complement the detailed profiling and analysis conducted in this research, especially where geodemographic profiles are created from sample data, and in the quest to uncover regularities in the determinants of participation rates.

Geodemographic analysis presents a statistical model which compresses the many characteristics of people living within areas into discrete categories. Throughout this thesis, the purpose of these Geodemographic representations has been to condense these multiple characteristics into measures which may easily be interpreted by end users. A major contribution of this thesis has been to explore how these summary models might be improved through addition of sector or application specific data. However, in completing this process, these models introduce additional uncertainty as they are based on the average characteristics of people, direct measures of a range of characteristics attributable to individuals. An alternative to the geodemographic approach is to represent the complex multivariate nature of a given dataset using more traditional multiple regression type analyses: however, for the audience for which much of this research was intended, such models were deemed much less readily interpretable.

In concluding, it is perhaps appropriate to identify some of the themes which will be carried forward by the author of this thesis in further post doctoral research. These include:

- The use of geodemographics to investigate the effects of the increase in tuition fees upon widening participation in post 2004 UCAS data.
- Further investigation of patterns of applicant behaviour in order to predict the extent to which applications cluster spatially or by attribute, in order to create behavioural models which could be used to guide applicants into appropriate or alternative course and institution choices.
- Extending work with UCAS to help develop the profile of the tools created in this thesis to enable them to better embed these research into their data outreach strategies and analytical services.

More generally, further work is also needed to integrate Geodemographic models with explanatory multivariate statistical modelling, specifically but not exclusively to better account for local spatial variation. Thus, one major research agenda unfolding from this thesis will be to assess how the spatial data infrastructure of the UK education sector might be focused through Geodemographics into the development of improved (and predictive) spatial interaction and spatial choice models that are sensitive to local context. This agenda extends the remit of this research from the realm of descriptive interpretation towards modelling with wider explanatory and predictive goals. It is also hoped that this may be achieved alongside the development of new tools that not only engage with new areas of policy but also provide information in ways that decision makers find easy to interpret and understand.